UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

KOKI FERNANDO OIKAWA

**Análise estatística de dados longitudinais e hierárquicos em Psicologia: uma
análise comparativa entre GEE e GLMM**

São Paulo
2019

KOKI FERNANDO OIKAWA

**Análise estatística de dados longitudinais e hierárquicos em Psicologia: uma
análise comparativa entre GEE e GLMM**

Tese apresentada ao Instituto de Psicologia da
Universidade de São Paulo como requisito parcial
para obtenção do título de Doutor em Psicologia.

Área de concentração: Psicologia Experimental

Orientador: Prof. Dr. José de Oliveira Siqueira

São Paulo

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na publicação
Serviço de Biblioteca e Documentação
Instituto de Psicologia da Universidade de São Paulo

Oikawa, Koki Fernando

Análise estatística de dados longitudinais e hierárquicos em Psicologia: uma análise comparativa entre GEE e GLMM / Koki Fernando Oikawa ; Orientador, José de Oliveira Siqueira. – 2019

150 f.

Tese (Doutorado em Psicologia) – Programa de Pós-Graduação em Psicologia Experimental, Instituto de Psicologia, Universidade de São Paulo, 2019.

Versão original

1. Modelos de Regressão. 2. Dados longitudinais. 3. Modelo GEE. 4. Modelo GLMM. 5. Modelo Population-Averaged (PA). 6. Modelo Subject-Specific (SS). 7. Simulação. I. Siqueira, José de Oliveira, orient, II. Título.

Nome: OIKAWA, Koki Fernando

Título: Análise estatística de dados longitudinais e hierárquicos em Psicologia: uma análise comparativa entre GEE e GLMM.

Tese apresentada ao Instituto de Psicologia da
Universidade de São Paulo como requisito parcial
para obtenção do título de Doutor em Psicologia.

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

AGRADECIMENTOS

Inicialmente gostaria de agradecer ao meu orientador, Prof. Dr. Siqueira, por sua temperança, equilíbrio e paciência comigo. Além disso, também sou muito agradecido ao senhor por ter me apresentado a um tema que só com o tempo fui entendendo algumas de suas mensagens iniciais: importante, interessante e pouco estudado (ao menos no Brasil).

Agradeço, também, à CAPES pelo financiamento da pesquisa.

Poderia ficar lembrando e agradecendo vários amigos e familiares, mas, desta vez, gostaria de fazer algo diferente. Pelo fato de ainda não ter tido a oportunidade de conhecê-los pessoalmente, uma vez que eles moram no Japão, dedico a minha Tese especialmente aos meus dois sobrinhos queridos que sempre estiveram em meus pensamentos: Raquel Megume Oikawa e Rafael Eidi Oikawa.

RESUMO

OIKAWA, Koki Fernando. (2019). **Análise estatística de dados longitudinais e hierárquicos em Psicologia: uma análise comparativa entre GEE e GLMM**. 2019. 213 f. Tese (Doutorado em Psicologia Experimental) – Instituto de Psicologia da Universidade de São Paulo, São Paulo, 2019.

Bancos de dados contendo características longitudinais surgem frequentemente na área de saúde. Dados longitudinais no nível da unidade experimental (pacientes, indivíduos em geral), podem ser considerados como observações sob a forma de medidas repetidas e em uma determinada ordem temporal. Em um contexto de dados longitudinais, medidas de associação e ou de correlação das respostas (das unidades experimentais) também devem ser levadas em consideração, uma vez que certos padrões de respostas (no nível das unidades experimentais) surgem regularmente. Usualmente, dois modelos são empregados para análise de dados longitudinais: a) GEE e b) GLMM. O primeiro é utilizado quando as questões de interesse residem sobre aspectos populacionais, enquanto que o segundo, para análises mais direcionadas mais especificamente no nível das unidades experimentais. Na década de 80, esses modelos passaram a ser considerados como estratégia para a análise de dados longitudinais tendo, por um lado, a classe de modelos PA (Population-Averaged) e, por outro, a classe de modelos SS (Subject-Specific). Mais recentemente, foi descoberto que essas classes de modelos possuem certas relações matemáticas de equivalência. Nessa Tese, serão realizadas simulações de dados longitudinais por meio de um pacote do software R especialmente desenvolvido para essa finalidade, com uma quantidade de unidades experimentais relativamente elevada ($N = 10.000$) e 6 medidas repetidas em cada uma delas, características pouco frequentes na literatura. Além disso, modelos das classes PA e SS serão estimados a partir dos dados simulados e o grau de aproximação das estimativas dos parâmetros (interceptos e coeficientes) será averiguada.

Palavras-chave: Modelos de Regressão, Dados longitudinais, Modelo GEE, Modelo GLMM, Modelo Population-Averaged (PA), Modelo Subject-Specific (SS).

ABSTRACT

OIKAWA, Koki Fernando. (2019). **Statistical analysis of longitudinal and hierarchical data in Psychology: A comparative analysis between GEE and GLMM**. 2019. 213 f. Tese (Doutorado em Psicologia Experimental) – Instituto de Psicologia da Universidade de São Paulo, São Paulo, 2019.

Dataset containing longitudinal characteristics often appear in the health area. Longitudinal data at the level of the experimental unit (patients, individuals, in general), can be considered as observations in the form of repeated measures and in a given temporal order. In a context of longitudinal dataset, measures of association and / or correlation of responses should also be taken into account, since response patterns can regularly emerge (at the level of experimental unit). Usually, two models are used for longitudinal data analysis: a) GEE and b) GLMM. The first one is used when populational aspects are of particular interest, while, the second one, for analysis at the experimental unit level. In the 1980s, both models were considered as part of a strategy for longitudinal data analysis with PA (Population-Averaged) models, on the one hand, and SS (Subject-Specific) models, on the other. More recently, it has been found that these kinds of models have certain mathematical equivalence relations. In this Thesis, longitudinal data simulations will be carried out by a specially R software package designed for this purpose, with a relatively high size of experimental units (N=10,000) and 6 repeated measures in each one of them, which is not very often in the literature. In addition, models from both PA and SS classes will be estimated from the simulated datasets and the degree of approximation / equivalence of the parameter estimates (intercepts and coefficients) will be examined.

Keywords: Regression Models, Longitudinal Dataset, GEE Model, GLMM Model, PA (Population-Averaged) Model, SS (Subject-Specific) Model.

LISTA DE FIGURAS

Figura 1.1 – Tendência das técnicas estatísticas (em percentual) em jornais e revistas de Psicologia de 1948 a 2001	3
Figura 5.1 – Tela HWLM	82
Figura 5.2 – Tela HWLM: Determinação do primeiro nível hierárquico	83
Figura 5.3 – Tela HWLM: Exemplo de modelo hierárquico de dois níveis	83
Figura 5.4 – Tela HWLM da saída do modelo Nulo	89
Figura 5.5 – Tela HWLM da saída do modelo 2	94
Figura 5.6 – Tela HWLM da saída do modelo 3	97
Figura 5.7 – Tela HWLM da saída do modelo 4	102
Figura 5.8 – Tela HWLM da saída do modelo 5	106
Figura 6.1 – Média marginal μ_{it} vs. η_{it} , com Y satisfazendo $\text{logit}EY b = \eta + b$	131

LISTA DE TABELAS

Tabela 1.1 – Percentuais médios com Intervalo de Confiança de 95%	4
Tabela 4.1 – Modelos, estimadores e algoritmos computacionais	75
Tabela 5.1 – Saídas das estimativas para o modelo Nulo.....	88
Tabela 5.2 – Estimativas dos ICCs para o modelo Nulo	89
Tabela 5.3 – Comando executado pelo R para o modelo Nulo	90
Tabela 5.4 – Comandos executados pelo SPSS para o modelo Nulo.....	90
Tabela 5.5 – Análise comparativa entre as tendências linear e quadrática (continua)	91
Tabela 5.6 – Saídas das estimativas para o modelo 2.....	93
Tabela 5.7 – Comando executado pelo R para o modelo 2	93
Tabela 5.8 – Comandos executados pelo SPSS para o modelo 2.....	93
Tabela 5.9 – Saídas das estimativas para o modelo 3.....	96
Tabela 5.10 – Comando executado pelo R para o modelo 3	96
Tabela 5.11 – Comandos executados pelo SPSS para o modelo 3.....	97
Tabela 5.12 – Saídas das estimativas para o modelo 4 (continua)	99
Tabela 5.13 – Comando executado pelo R para o modelo 4	100
Tabela 5.14 – Comandos executados pelo SPSS para o modelo 4.....	100
Tabela 5.15 – Saídas das estimativas para o modelo 5.....	103
Tabela 5.16 – Comando executado pelo R para o modelo 4	104
Tabela 5.17 – Comandos executados pelo SPSS para o modelo 5.....	104
Tabela 6.1 – Relação entre média condicional, função de ligação e média marginal para classe de respostas GLM.....	132
Tabela 6.2 – Resultados do Modelo PA – Harvard Study of Air Pollution and Health	133
Tabela 6.3 – Resultados do Modelo SS – <i>Harvard Study of Air Pollution and Health</i>	134
Tabela 6.4 – Estimativas dos modelos PA e SS para dados de doenças nos seios.....	138
Tabela 6.5 – Estimativas de PA e SS para modelo Pós-Teste	146
Tabela 6.6 – Estimativas das correlações nos modelos PA e SS.....	147
Tabela 6.7 – Estimativas de modelos PA	149
Tabela 6.8 – Estimativas de modelos SS.....	150
Tabela 6.9 – Relação entre os parâmetros dos modelos PA e SS.....	151
Tabela 6.10 – Correlações intraclasse entre modelos GEE e GLMM.....	153
Tabela 6.11 – Estimativas GEE e GLMM para três situações específicas de zi	159
Tabela 6.12 – Estimativas de parâmetros e de erros tipo I sob condições específicas	160

Tabela 6.13 – Comparação entre os métodos de estimação MLE e SVE	162
Tabela 6.14 – Estimativas dos modelos GEE e GLMM de dados de alcoolismo	163
Tabela 7.1 – Grupos de Análise.....	174
Tabela 7.2 – Estimativas dos modelos SS (regressão de Poisson e função de ligação <i>log</i>) ...	176
Tabela 7.3 – Estimativas dos modelos PA (regressão de Poisson e função de ligação <i>log</i>) ..	177
Tabela 7.4 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão de Poisson com função de ligação <i>log</i>	178
Tabela 7.5 – Comandos do R para modelos PA e SS de Regressão de Poisson com função de ligação <i>log</i>	178
Tabela 7.6 – Estimativas dos modelos SS (regressão logística e função de ligação <i>logit</i>)	180
Tabela 7.7 – Estimativas dos modelos PA (regressão logística e função de ligação <i>logit</i>)....	180
Tabela 7.8 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão logística com função de ligação <i>logit</i>	182
Tabela 7.9 – Comandos do R para modelos PA e SS de Regressão Logística com função de ligação <i>logit</i>	182
Tabela 7.10 – Estimativas dos modelos SS (regressão logística e função de ligação <i>probit</i>)	185
Tabela 7.11 – Estimativas dos modelos PA (regressão logística e função de ligação <i>probit</i>)	185
Tabela 7.12 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão logística com função de ligação <i>probit</i>	186
Tabela 7.13 - Comandos do R para modelos PA e SS de Regressão Logística com função de ligação <i>probit</i>	187

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 A estatística na Psicologia	1
1.2 O problema dos dados longitudinais.....	6
1.3 Modelos para dados longitudinais	8
2 O Modelo GEE	12
2.1 Introdução	12
2.2 A origem do GEE	12
2.3 O modelo logístico longitudinal	13
2.4 O modelo GEE com função de verossimilhança	18
2.5 O modelo GEE com função de quase-verossimilhança.....	24
2.6 Apêndice	30
3 O Modelo GLMM	36
3.1 Introdução	36
3.2 O modelo LMM	37
3.2.1 Estimação e Inferência	40
3.2.2 Usando o Algoritmo E-M para Estimativas de Máxima Verossimilhança	43
3.2.3 Estimação REML e Cálculo	44
3.3 O modelo GLMM	46
3.3.1 A Função de Quase-verossimilhança	48
3.3.2 O método PQL	51
3.3.3 Algoritmo de Estimação.....	53
3.3.4 Componentes de Variância.....	54
3.4 Apêndice	56
4 O Modelo Linear Hierárquico Generalizado (HGLM).....	64
4.1 Introdução	64
4.2 O modelo HLM.....	64
4.2.1 O Modelo HLM em Dois Níveis.....	65
4.2.2 Estimação dos Componentes de Variância	68
4.3 O modelo HGLM.....	71

4.3.1 O Modelo HGLM para Respostas Binárias.....	72
4.3.2 O Modelo HGLM para Respostas de Poisson.....	73
4.3.3 Métodos de Estimação para HGLM.....	74
4.4 A relação entre HGLM e GLMM	76
5 O Software HLM 7	79
5.1 Introdução	79
5.2 Características do software HLM 7	79
5.3 Análise de um banco de dados	85
5.3.1 Modelo 1: Modelo Nulo	87
5.3.2 Modelo 2: Adicionando Taxas de Crescimento	91
5.3.3 Modelo 3: Adicionando <i>time-varying covariates</i>	95
5.3.4 Modelo 4: Explorando Diferenças nas Trajetórias de Crescimento entre Instituições	98
5.3.5 Modelo 5: Examinar as Taxas de Crescimento no Nível 3	103
5.4 Apêndice	108
6 Algumas relações entre os modelos PA e SS	126
6.1 Introdução	126
6.2 Os modelos PA e SS – Regressão Logística.....	126
6.2.1 A Proposta de Zeger, Liang e Albert.....	126
6.2.2 Os Modelos PA e SS – Alguns Resultados Gerais para Regressão Logística	135
6.3 Os modelos PA e SS – Regressão de Poisson (parte I)	139
6.3.1 A Análise do Modelo Pós-teste.....	144
6.3.2 A Análise do Modelo Pré-teste/Pós-teste.....	147
6.4 Os modelos PA e SS – Regressão de Poisson (parte II)	153
6.4.1 Análise de Dados Simulados.....	157
6.4.2 Análise de Dados Reais.....	162
6.5 Modelos PA e SS – Algumas relações gerais	164
6.6 Apêndice	168
7 Simulações e análises comparativas entre os modelos PA e SS.....	170
7.1 Introdução	170

7.2 As estratégias PA e SS.....	170
7.2.1 Estratégia PA e SS: o Modelo de Regressão de Poisson	171
7.2.2 Estratégia PA e SS: o Modelo de Regressão Logística com Função de Ligação <i>logit</i>	171
7.2.3 Relação entre os Modelos SS e PA: Regressão Logística com Função de Ligação <i>probit</i>	172
7.3 Simulações e comparações dos parâmetros PA e SS.....	173
7.3.1 Relação entre PA-SS via Regressão de Poisson com Função de Ligação <i>log</i>	175
7.3.2 Relação entre PA e SS via Regressão Logística com Função de Ligação <i>logit</i>	179
7.3.3 Relação entre PA e SS via Regressão Logística com Função de Ligação <i>probit</i> ..	184
7.4 Conclusões	189
8 CONCLUSÕES E LIMITAÇÕES	191
9 REFERÊNCIAS	193

1 INTRODUÇÃO

1.1 A estatística na Psicologia

Os profissionais da Psicologia vêm empreendendo, já há algum tempo, esforços para elevar o volume de pesquisas com o apoio de métodos quantitativos. O Conselho de Representantes da APA (*American Psychological Association*) autorizou, em 2006, uma força-tarefa para aumentar o número de psicólogos quantitativos, a qual produziu, em 2009, o relatório “Report of the Task Force for Increasing the Number of Quantitative Psychologists” (Aiken et al., 2009).

No relatório, a comissão responsável faz uma análise da escassez de psicólogos quantitativos, do mercado de trabalho para esse tipo de profissional e da necessidade de mais profissionais nessa área em um contexto de Psicologia Científica.

Na primeira página do relatório (Aiken et al., 2009), quando da elaboração de uma definição mais precisa acerca dessa disciplina, a Estatística é citada de maneira tal que se encontra intrinsecamente associada à própria significação delimitada na pesquisa: “*Quantitative psychology is the study of methods and techniques for the measurement of human attributes, the statistical and mathematical modeling of psychological processes, the design of research studies, and the analysis of psychological data*”. (p. 1)

Um dos questionamentos de interesse da comissão (Questão 2) estava relacionado ao problema de como elevar o número de estudantes (com certo *background*) em Programas de Doutorado em Psicologia Quantitativa. Além disso, o relatório também destacava medidas já adotadas pela comissão responsável, sendo algumas para criação de informações acessíveis (de Psicologia Quantitativa) e outras específicas para o desenvolvimento de relações cooperativas com estruturas organizacionais já existentes. Por fim, o relatório apresentou um conjunto de recomendações no intuito de alcançar os resultados almejados.

Esse tipo de posicionamento/atitude por parte da APA foi levado em consideração em Cousineau (2005), em um estudo sobre o surgimento dos métodos quantitativos na Psicologia: “*Here, we examine the question whether psychology is ready to enter the ‘hard science club’, like biology did in the forties.*” (p. ??)

Existem vários estudos sobre o papel que os métodos e técnicas estatísticas têm exercido sobre a área de Psicologia. Skidmore and Thompson (2010), por exemplo, fizeram um levantamento sobre as técnicas estatísticas mais comumente empregadas nas áreas da Educação

e da Psicologia. Nesse estudo, 12.012 artigos serviram de base para análise, a partir dos quais foram registrados, em ambas as áreas, um total de 17.698 aplicações de técnicas estatísticas.

Esse tipo de estudo merece destaque na medida em que permite avaliar os questionamentos científicos que têm sido pautados pelos psicólogos profissionais, bem como as estratégias escolhidas na busca de solução(ões) desses questionamentos. Esse tipo de análise permite uma investigação mais meticulosa sobre a natureza dos próprios questionamentos científicos, desde a escolha do tipo de experimento a ser implementado até a determinação da(s) técnica(s) estatística(s) para a análise dos dados obtidos.

Skidmore e Thompson (2010) observaram que Edgington (1964) já sinalizava para a importância de se identificar, dentro da literatura estatística, os conhecimentos necessários aos estudantes de doutorado em Psicologia para que fossem capazes de avaliar o que poderia vir a ser útil.

Nessa pesquisa conduzida pelos autores (Skidmore & Thompson, 2010), artigos com apenas uma única técnica estatística foram excluídos e duas foram as razões alegadas pelos autores: i) utilizar uma única técnica estatística não estaria de acordo com a realidade vigente, isto é, a grande maioria dos artigos utilizava pelo menos duas técnicas estatísticas, e ii) tal estratégia (analisar artigos com pelo menos duas técnicas estatísticas) tornaria comparável todos os artigos, então, considerados. Essa estratégia de seleção dos artigos foi aplicada em ambas as áreas (Educação e Psicologia).

Especificamente com relação aos artigos de Psicologia, foram considerados os seguintes periódicos científicos selecionados pela APA e pela *American Counseling Association* (ACA) entre 1948 e 2001: *Journal of Counseling and Development* (JCD), *Journal of Educational Psychology* (JEP), *Journal of Counseling Psychology* (JCP) e *Journal of Personality Assessment* (JPA).

Na Figura 1.1, encontram-se seis gráficos referentes a seis grupos de técnicas estatísticas mais empregadas nos artigos de Psicologia publicados nos periódicos mencionados acima, dentro do período considerado. Esses seis grupos de técnicas estatísticas são:

- a) métodos não paramétricos;
- b) teste t;
- c) correlação;
- d) regressão;
- e) ANOVA/ANCOVA;
- f) análise de cluster e de fator.

Por meio da Figura 1.1, é possível observar que a frequência da utilização de *técnicas não paramétricas* tem variado ao longo do tempo, tendo exibido crescimento (até os anos de 1960), seguido de diminuição (até os anos de 1990) para, a partir de então, voltar a crescer.

Figura 1.1 – Tendência das técnicas estatísticas (em percentual) em jornais e revistas de Psicologia de 1948 a 2001

Fonte: Skidmore e Thompson (2010)

Com relação ao *teste t*, uma tendência de queda, de um modo geral, tem se revelado para o período considerado, com a observação de que, entre os anos de 1960 e 1990, essa tendência de queda foi interrompida.

O gráfico da frequência das técnicas de *correlação* tem exibido uma queda acentuada (até os anos de 1960), seguida de uma trajetória ascendente (até os anos de 1990), para, em seguida, voltar a declinar. Em particular, a frequência das técnicas de *correlação* apresentou um

comportamento contrário quando comparado com a frequência dos *métodos não paramétricos*, no sentido de que a tendência de crescimento (ou decrescimento) de uma técnica, em geral, condiz com o decrescimento (ou crescimento) da outra.

A frequência das técnicas de *regressão* apresentou uma tendência geral de crescimento, tendo esboçado, entre os anos de 1980 e 1995, uma leve tendência a contração.

A aplicação das técnicas ANOVA/ANCOVA cresceu até os de 1970 (ponto de máximo) para, então, diminuir de forma continuada e sistemática.

Por fim, o grupo das técnicas de *análises de cluster e fator*, até os anos de 1960, exibiu uma tendência relativamente estacionária e, em seguida, passou a apresentar uma trajetória crescente. Skidmore e Thompson (2010) relatam que a matemática envolvida em ambas as técnicas seria intimidadora ao ponto de inibir aplicações, sendo que, somente após a disseminação do uso de softwares estatísticos, essas técnicas puderam ser mais utilizadas. A Tabela 1.1, também extraída de Skidmore e Thompson (2010), apresenta um resumo para um período mais recente (década de 1990) acerca do percentual médio (M) da frequência utilizada das referidas técnicas estatísticas e do desvio padrão estimados pelos próprios autores. Esses percentuais médios são ordenados por artigo, entre os anos de 1990 e 1997.

Tabela 1.1 – Percentuais médios com Intervalo de Confiança de 95%

Técnica estatística	Psicologia (n=10)			
	95% IC			
	M	SD	LL	UL
Não Paramétrica	18,38	7,93	12,71	24,05
Teste t	7,73	10,16	0,45	15,00
Correlação	46,60	12,76	37,47	55,73
Regressão	25,80	6,70	21,00	30,59
ANOVA/ANCOVA	30,71	13,60	20,99	40,44
Análise de Fator/Cluster	15,58	8,01	9,85	21,31

Fonte: Skidmore e Thompson (2010)

É possível notar, ainda, que a soma das proporções não resulta em 100%, uma vez que, em cada artigo considerado, duas ou mais técnicas foram utilizadas.

O resumo apresentado na Tabela 1.1 refere-se a uma extração (recorte) de uma tabela mais geral (Educação, Psicologia e todos os jornais) do texto original de Skidmore e Thompson

(2010). Uma vez que o interesse reside no estudo das aplicações das técnicas estatísticas na Psicologia, apenas essa dimensão dos dados foi extraída.

De acordo, ainda, com os dados da Tabela 1.1, as três técnicas estatísticas mais utilizadas são correlação (46,6%), aparecendo em quase metade dos artigos, seguida da ANOVA/ANCOVA (30,7%) e da regressão (25,8%). As colunas LL e UL representam os limites inferior e superior, respectivamente, dos intervalos de confiança para os percentuais (médios) de cada técnica estatística, com 95% de confiança.

Com relação ao declínio do emprego do teste *t* e aumento das aplicações da técnica ANOVA até início da década de 1970 (Figura 1.1), Skidmore e Thompson (2010) chamam atenção para uma falha por parte dos pesquisadores de Psicologia e Educação. Segundo Skidmore e Thompson (2010), os profissionais dessas áreas têm, historicamente, considerado que os experimentos constituem o único tipo de delineamento capaz de solucionar problemas de causalidade. Em seguida, há uma menção de Thompson (2006, p. 386), a saber, de que os pesquisadores dessas áreas detinham o conhecimento de que a técnica ANOVA seria útil quando variáveis independentes fossem nominais e as variáveis dependentes, intervalares e que, sendo assim, eles começaram a associar “inconscientemente” a técnica ANOVA com o planejamento de um experimento. Mas, conforme observação de Skidmore e Thompson (2010), o problema não estaria precisamente nessa associação, mas, sim, na pressuposição de que relações causais poderiam ser obtidas a partir da aplicação da técnica ANOVA.

Com relação à tendência crescente de utilização de técnicas de Regressão, os autores apontam para Graham (2008), o qual alega que:

The vast majority of parametric statistical procedures in common use are part of [a single analytic family called] the General Linear Model (GLM), including the t test, analysis of variance (ANOVA) multiple regression, descriptive discriminant analysis (DDA), multivariate analysis of variance (MANOVA), canonical correlation analysis (CCA), and structural equation modeling (SEM). Moreover, these procedures are hierarchical [italics added], in that some procedures are special cases of others. (p. 485)

Aqui, é importante ressaltar que o autor faz referência ao GLM, no sentido de modelo linear geral que, na realidade, é relativamente diferente do sentido que será empregado no decorrer da tese. A expressão GLM empregada por Graham diz respeito, fundamentalmente, à estrutura geral dos modelos lineares que, apesar de estarem relacionados com o tema da tese, não será, a princípio, o assunto central.

Contudo, o termo GLM a ser amplamente utilizado no decorrer desta tese diz respeito aos modelos lineares generalizados (*Generalized Linear Models*), a saber, uma classe especial de modelos de regressão que permite que uma variável resposta possa assumir distribuição de

probabilidades diferente da normal. Por exemplo, variável resposta com distribuição binomial (regressão logística), Poisson (regressão de Poisson) etc. Assim, a classe de modelos GLM a ser utilizada no decorrer desta tese é a *Generalized Linear Models* (e não *General Linear Models*).

A seguir, será feita uma breve e sucinta exposição sobre essas técnicas estatísticas tão amplamente exploradas na Psicologia, contidas na Tabela 1.1.

A **correlação**, de um modo geral, consiste em uma medida de associação entre variáveis, podendo, ainda, descrever a tendência dessa associação, isto é, se negativa, se positiva, forte ou fraca. Existem várias medidas de correlação e a literatura estatística é vasta.

É possível, no entanto, que o interesse não esteja somente no grau de associação entre variáveis, mas em saber se existe diferença significativa entre elas. Uma técnica estatística capaz de realizar esse tipo de análise é o **teste t**, utilizada frequentemente para testar a diferença entre grupos.

Um outro tipo de comparação de mais de dois grupos pode ser feito por meio da análise de variância, **ANOVA**. É interessante ressaltar que existem modelos ANOVA de variados tipos, apropriados para diversas situações específicas. Mas a ideia fundamental continua sendo a comparação de mais de dois grupos.

As técnicas de **regressão** permitem medir, avaliar e até mesmo prever o efeito que uma variável explicativa pode exercer em uma variável dependente. Assim, enquanto a correlação avalia o quão fortemente associadas duas variáveis aleatórias podem ser (em magnitude e direção), a regressão permite estimar o quanto uma variável dependente mudaria, caso a variável explicativa também se alterasse.

A análise de **fatores** consiste em um método capaz de descrever a variabilidade entre variáveis observáveis, eventualmente correlacionadas, em termos de outras variáveis não observáveis, denominadas fatores, geralmente em menor número. Nesse tipo de análise, grupos de variáveis fortemente correlacionadas podem formar um fator, o qual pode ser considerado como variável latente.

1.2 O problema dos dados longitudinais

Fitzmaurice e Molenberghs (2009) colocaram em uma perspectiva histórica o problema da análise de dados longitudinais. Nesta seção, serão abordados alguns dos apontamentos feitos por esses autores (Fitzmaurice & Molenberghs, 2009) que tratam das limitações que certas técnicas estatísticas possuem quando aplicadas em dados longitudinais.

Fitzmaurice e Molenberghs (2009) destacam que, em um contexto longitudinal, a técnica estatística ANOVA tem exercido um papel relevante na análise da(s) mudança(s).

Segundo, ainda, Fitzmaurice e Molenberghs (2009), a técnica ANOVA de efeitos mistos (*mixed-effects ANOVA*), isto é, a técnica ANOVA com a introdução de um efeito aleatório por UE, tem sido historicamente empregada para a análise de dados longitudinais.

A técnica ANOVA de efeitos mistos também é conhecida como ANOVA de medidas repetidas univariada, e, em um contexto de dados longitudinais, as UEs são consideradas como blocos cujos efeitos são aleatórios. Em particular, esses efeitos aleatórios constituem uma conjunção de fatores (não observáveis e não mensuráveis) que influenciam a própria diferenciação das respostas das UEs, sendo que a simples introdução desses efeitos aleatórios implica na indução de correlação positiva das respostas.

Fitzmaurice e Molenberghs (2009) ressaltam que, se por um lado, a ANOVA de medidas repetidas univariada pode ser considerada como uma espécie de generalização do teste t pareado ao ser capaz de lidar com mais de duas mensurações, por outro, o fato do efeito dos blocos ser tratado como aleatório (e não como fixo) constitui um problema na medida em que a aleatorização em cada um dos períodos e ao longo do tempo passa a ser um argumento sem sentido.

No entanto, Fitzmaurice e Molenberghs (2009) observaram que, apesar dessa restrição, a ANOVA de medidas repetidas univariada pode ser considerada como uma espécie de precursora de modelos para dados longitudinais.

Ainda nessa perspectiva histórica, uma versão multivariada da ANOVA de medidas repetidas (MANOVA de medidas repetidas multivariada) foi desenvolvida, permitindo analisar um vetor de variáveis respostas distintas.

Contudo, conforme analisaram Fitzmaurice e Molenberghs (2009), a técnica MANOVA de medidas repetidas multivariada ainda não constitui uma técnica apropriada para o caso dos dados longitudinais. Enquanto que os dados longitudinais podem ser analisados como um vetor de uma mesma variável resposta, a técnica MANOVA de medidas repetidas multivariada foi originalmente desenvolvida para a análise de vetor formado por múltiplas variáveis respostas.

Além disso, Fitzmaurice e Molenberghs (2009) notaram duas características que particularmente limitam a utilidade da MANOVA em um contexto de dados longitudinais: i) quando os experimentos são desbalanceados e ii) presença de padrões gerais de dados faltantes (ao menos em alguns softwares estatísticos).

No entanto, como é possível haver, em estudos longitudinais, dados desbalanceados nas UEs (seja em razão dos números de observações, seja em razão da existência de distintas

sequências de tempo), as especificidades bem como as características que são próprias desse tipo de estrutura de dados têm motivado o desenvolvimento de novas classes de modelos que sejam capazes de lidar com essas peculiaridades.

1.3 Modelos para dados longitudinais

Nas décadas de 1980 e 1990, duas novas classes de modelos de regressão foram desenvolvidas para lidar com a estrutura de dados longitudinais:

- i) GEE (*Generalized Estimating Equations*);
- ii) GLMM (*Generalized Linear Mixed Models*).

Essas classes de modelos representam um salto de qualidade para a modelagem de dados relacionados aos problemas científicos, em geral, e da área da saúde, em particular, tendo ambas sido originadas em departamentos de Bioestatística: o GEE, pela Johns Hopkins, e o GLMM, pela Harvard.

A classe de modelos GEE, por exemplo, permite estimar parâmetros de um modelo de regressão em um contexto de dados longitudinais, levando em consideração a correlação entre as respostas das UEs.

Contudo, a classe de modelos GLMM também permite estimar, em um contexto de dados longitudinais, parâmetros de um modelo de regressão com correlação entre as respostas das UEs. Existem várias diferenças de ordem metodológica que serão analisadas em detalhes nos capítulos subsequentes, mas cabe aqui destacar que a forma como a classe de modelos GLMM leva em consideração a correlação entre as respostas dá-se via introdução de efeitos aleatórios (por meio de interceptos aleatórios), em nível individual (para cada UE), expressando, dessa maneira, a heterogeneidade natural de cada UE, devido a fatores não observáveis e ou não mensuráveis.

Uma classe anterior de modelos com efeitos aleatórios inicialmente desenvolvida por Laird e Ware (1982) considerava apenas o caso de respostas com distribuição normal, conhecida como Modelos Mistos Lineares (*Linear Mixed Models* – LMM).

Breslow e Clayton (1993) empreendem uma valiosa contribuição ao desenvolver a classe GLMM de modelos (também com efeitos aleatórios), mas com a possibilidade das variáveis respostas serem contínuas ou discretas, constituindo uma extensão da classe de modelos GLM (*Generalized Linear Models*), desenvolvida por Nelder e Wedderburn (1972). Logo, a classe de modelos GLMM estende não somente a classe LMM ao permitir outras distribuições de probabilidade para a variável resposta, mas também a classe GLM, ao levar em consideração

efeitos aleatórios para cada UE, sem perder de vista que a base de dados em questão conta com uma estrutura longitudinal, com a possibilidade de se levar em consideração a correlação das respostas.

É importante notar que as classes de modelos GEE e GLMM também são apropriadas para lidar com problemas de dados de agrupamento (*cluster*), não se restringindo, portanto, apenas para o caso de problemas (e dados) longitudinais.

No caso particular da classe de modelos GLMM, é conhecida sua relação com a classe de modelos hierárquicos (ou multiníveis) (Raudenbush & Bryk, 2002, p. 292), fato a ser analisado no Capítulo 4.

Contudo, uma relação entre os parâmetros dos próprios modelos GLMM e GEE também é conhecida em um contexto de dados longitudinais. Nesse caso, Zeger, Liang e Albert (1988) consideram duas estratégias para a análise de dados longitudinais: i) modelos PA (*Population-Averaged*) e ii) modelos SS (*Subject-Specific*).

A classe de modelos PA estaria voltada para estimação e interpretação de parâmetros em uma perspectiva populacional, enquanto que, para a classe de modelos SS, o foco das análises estaria no nível das UEs. Se a abordagem fosse por modelos PA, a escolha deveria ser pela classe de modelos GEE, enquanto que, se a abordagem fosse para análises no nível das UEs (SS), a escolha deveria ser para a classe de modelos GLMM.

Assim, partindo da problemática do Relatório da APA, isto é, da necessidade de se elevar a taxa de psicólogos quantitativos, bem como dos levantamentos de Skidmore e Thompson (2010), é necessário destacar a existência (bem como os desenvolvimentos recentes) de classes de modelos próprios e específicos para a análise de dados longitudinais.

Além disso, uma análise mais pormenorizada acerca da existência de uma relação de equivalência entre os parâmetros desses modelos (GEE e GLMM), bem como das condicionantes dessa relação, será realizada no intuito de obter um conhecimento mais detalhado tanto dessas classes de modelos quanto da estratégia para abordagem de análise de dados longitudinais originalmente proposta por Zeger et al. (1988).

Assim, ao analisar uma base de dados longitudinais, quais opções de modelos (e estratégias possíveis) estariam dispostas para um psicólogo quantitativo? Como funcionam essas classes de modelos? Que interpretações paramétricas são possíveis em cada um dos casos? Existem relações matemáticas aproximadas entre os parâmetros dos modelos GEE e GLMM (ou seja, entre as classes de modelos PA e SS)? Em quais condições essas relações seriam válidas?

Além disso, nesta tese será introduzido um pacote do software R (*simstudy*) especializado em simulações de bases de dados longitudinais, e bases de dados com 10.000 UEs serão simuladas (todas com seis períodos de observação ao longo do tempo), tamanho amostral relativamente incomum na literatura. Acredita-se que essa seria uma oportunidade para verificar essas relações entre os modelos PA e SS em circunstâncias variadas (modelos de regressão logística e de Poisson com distintas funções de ligação e estruturas de correlação), com dados simulados e um número elevado de tamanho amostral.

De maneira a alcançar esses objetivos, esta tese encontra-se estruturada da seguinte maneira:

No segundo capítulo, será introduzida a classe de modelos GEE. Uma breve contextualização histórica será feita acerca do surgimento dessa classe de modelo.

No terceiro capítulo, será introduzida a classe de modelos GLMM, mas, antes, uma abordagem sobre a classe de modelos LMM (*Linear Mixed Models*) também será empreendida. Essa classe diferencia-se da GLMM por não ser *generalized*, isto é, admite apenas e tão somente a distribuição normal para a variável resposta (e não uma família de distribuições de probabilidade denominada família exponencial).

No quarto capítulo, será introduzida a classe de modelos lineares hierárquicos e uma base de dados com características hierárquica e longitudinal será analisada.

Para isso, será utilizado, no quinto capítulo, o software HLM 7, específico para análise de dados dessa área (hierárquicos), e uma comparação com outros softwares também será conduzida, no intuito de mostrar algumas das vantagens desse software, quando a análise é realizada a partir de uma perspectiva de um “não especialista” em softwares estatísticos renomados (R e ou SPSS).

No sexto capítulo, será desenvolvida a discussão sobre a relação entre duas estratégias para lidar com dados longitudinais, a saber, modelos PA e SS, sendo a primeira estimada pela classe de modelos GEE e a segunda estimada pela classe GLMM de modelos. Após a introdução desse tipo de estratégia, uma relação matemática aproximada entre os parâmetros dos modelos GEE e GLMM será analisada tanto para modelos de regressão logística quanto para modelos de regressão com resposta de Poisson.

No sétimo capítulo, será conduzida uma simulação de dados longitudinais por meio da biblioteca *simstudy* do software R, apropriada para esse tipo de atividade, no intuito de implementar as estratégias PA e SS e realizar comparações, bem como verificar as aproximações entre os parâmetros dos modelos GEE e GLMM. As bibliotecas *geepack* MASS e *nlme* também serão utilizadas, mas em um contexto de estimação de modelos.

Por fim, as conclusões gerais da tese, bem como sugestões para projetos futuros serão fornecidas no Capítulo 8.

2 O MODELO GEE

2.1 Introdução

Neste capítulo, será apresentada a classe de modelos GEE por meio de uma análise dos artigos que a originaram, sendo, o primeiro deles, denominado “Longitudinal data analysis using generalized linear models”, de Liang e Zeger, de 1986, e o segundo, de Zeger e Liang, denominado “Longitudinal Data Analysis for Discrete and Continuous Outcomes”, também de 1986. Um artigo anterior de Zeger, Liang e Self (1985), nem sempre tão comentado, também foi importante para o processo de elaboração da classe dos modelos GEE: “The Analysis of Binary Longitudinal Data with Time-Independent Covariates”. A análise desse artigo também receberá destaque neste capítulo.

2.2 A origem do GEE

Logo no primeiro parágrafo do livro *Generalized Estimating Equations*, Ziegler (2011) afirma que a modelagem GEE vem se tornando bastante popular nas áreas de Biometria, Econometria e Psicométrica por permitir superar certas condições bastante tradicionais da modelagem estatística, a saber, independência (entre observações) e normalidade, as quais podem ser, em muitas situações práticas, bastante restritivas. Além disso, enfatiza, também, que o desenvolvimento de técnicas de regressão tais como GEE foram propiciadas sobretudo pelo progresso computacional, de um modo geral, e pela construção de algoritmos iterativos, em particular.

Os especialistas da área costumam sugerir que dois foram os artigos responsáveis pelo surgimento da classe de modelos GEE (ambos publicados em revistas de Biometria). Esses seriam, seguindo a ordem cronológica, Zeger e Liang (1986) “Longitudinal Data Analysis for Discrete and Continuous Outcomes” e Liang e Zeger (1986) “Longitudinal data analysis using generalized linear models”, tendo sido publicados nos meses de março e abril, respectivamente. Ziegler (2011) sugere, ainda, mais um artigo responsável pela origem dos modelos GEE, a saber, o de Zeger et al. (1985) “The analysis of binary longitudinal data with time-independent covariates”, publicado um ano antes (também em uma revista de Biometria).

Contudo, uma menção deve ser feita com relação aos dois artigos de 1986. A ordem de publicação (cronológica) difere da própria ordem de citação entre os mesmos. O artigo de abril cita apenas o artigo de 1985 e não faz menção alguma ao artigo de março. Já o artigo de março

cita ambos (o artigo de abril e o de 1985), de modo que, aparentemente, apesar das datas de publicação, a ordem da elaboração dos artigos seja distinta, o que parece fazer sentido, uma vez que, no artigo de abril, os autores estenderam a família de modelos lineares generalizados (GLM) para a análise de dados longitudinais, enquanto que, no artigo de março, os autores acabam propondo uma abordagem unificadora para uma variedade de respostas (discretas e contínuas), fazendo uso da função de quase-verossimilhança, a qual exige apenas a especificação de uma relação entre a resposta média e as covariáveis. Vale lembrar que, na seção 1 do artigo de março, os autores enfatizam que:

The broad objectives of this paper are not to give details about quasi-likelihood and its extension to longitudinal data nor to give technical results about GEE. Rather, our intention is to motivate what we consider to be a widely applicable methodology for longitudinal data, summarize its advantages and disadvantages, and illustrate its use. (Zeger & Liang, 1986, p. 122)

Desse modo, analisaremos, a seguir, esses três artigos que criaram a classe de modelos GEE, em detalhes.

Inicialmente, será feita uma breve análise sobre o artigo de Zeger et al. (1985) “The analysis of binary longitudinal data with time-independent covariates”, dando destaque para o surgimento de um modelo para dados longitudinais, ainda que restrito ao contexto da regressão logística. Ziegler (2011) considera, em seu prefácio, esse artigo também como um dos artigos que deram origem à classe de modelos GEE. Em seguida, será analisado o artigo de Liang e Zeger (1986) “Longitudinal data analysis using generalized linear models” para, finalmente, discutir as implicações de Zeger e Liang (1986) “Longitudinal Data Analysis for Discrete and Continuous Outcomes”.

2.3 O modelo logístico longitudinal

No artigo de Zeger et al. (1985) “The analysis of binary longitudinal data with time-independent covariates”, os autores propuseram um modelo específico de regressão logística para uma situação na qual a variável resposta (binária) fosse passível de ser observada repetida no nível da UE. A obtenção de estimativas consistentes tanto dos coeficientes de regressão quanto de suas respectivas variâncias foi alcançada pelos autores ainda que sob certas suposições acerca da dependência temporal (*time dependence*) das trajetórias individuais. Uma aplicação em dados de estresse maternal de forma a exibir a técnica foi realizada.

Zeger et al. (1985) estendem o modelo de regressão logística para análise de dados longitudinais, mas de uma perspectiva particular. De certa forma, extensões dessa natureza, isto

é, de um modelo de regressão logística para dados longitudinais já haviam sido realizadas anteriormente, conforme relatado pelos próprios autores. No caso, Cox (1958), já havia sugerido um modelo (de regressão logística para dados longitudinais) no qual probabilidades condicionais de uma cadeia de Markov seriam expressas como funções logísticas de covariáveis. Para o caso em que as covariáveis fossem categóricas, uma possibilidade seria aquela proposta em Starmer e Koch (1969) e Koch, Landis, Freeman e Lehnen (1977).

Em contrapartida, se a resposta binária também fosse uma série temporal, então, métodos que levassem em consideração o processo de autocorrelação seriam necessários e é precisamente por esse caminho que os autores deixaram suas contribuições. Nesta seção serão introduzidas, de modo sucinto, as bases desse tipo de modelo de regressão logística, de modo que as respostas binárias de UEs possam ser analisadas como uma série temporal.

Do ponto de vista da notação, Zeger et al. (1985) consideraram Y_{it} ($t = 1, \dots, n_i$) uma série temporal estacionária e Z_i um vetor $s \times 1$ de covariáveis para a i -ésima UE, ($i = 1, \dots, K$), sendo K o número total de UEs. Note, nesse caso, que Z_i é apenas um vetor de covariáveis (e não uma matriz de todos os dados da i -ésima UE). Uma vez definido $\pi_i \equiv \mathbb{P}(Y_{it} = 1|Z_i)$, os autores assumiram que $\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = Z_i'\beta$.

É importante destacar que Zeger et al. (1985) estavam preocupados em modelar a probabilidade marginal (e não a probabilidade condicional) expressa como função das covariáveis e que, além disso, a abordagem foi especificamente de séries temporais (e não de medidas repetidas) para levar em consideração a dependência temporal entre as respostas binárias ao longo do tempo.

Para isso, os autores propõem dois estimadores para β : i) $\hat{\beta}_0$ e ii) $\hat{\beta}_1$.

O primeiro, $\hat{\beta}_0$, parte da suposição de que as respostas observadas em uma mesma UE são independentes entre si, enquanto que, por outro lado, o segundo estimador, $\hat{\beta}_1$, parte do pressuposto de que cada série constitui uma cadeia de Markov estacionária de ordem 1.

O primeiro estimador é consistente para quaisquer conjuntos de processos binários estacionários, enquanto que a consistência do segundo estimador também estaria satisfeita na medida em que as séries: i) satisfaçam a suposição da função de ligação *logit* e ii) admitam um processo de autocorrelação de ordem 1 em comum, dado por $\rho = \text{Corr}(Y_{it}, Y_{i,t-1})$.

Sob a suposição de independência, a análise de verossimilhança tradicional (seja para estimação, seja para inferência) seria bastante adequada. Seja a equação de escore $S_u(\beta)$, usualmente empregada para determinar os valores de β_u ($u = 1, \dots, s$), e $I_0(\beta)$, a matriz da

informação de Fisher. Com base nessas premissas, Zeger et al. (1985) sugerem a seguinte proposição:

Proposição 1. Seja Y_{it} ($t = 1, \dots, n_i < \infty; i = 1, \dots, K$) uma série binária estacionária tal que $\text{logit } \mathbb{E}(Y_{it}|Z'_i) = Z'_i\beta$. Então, $K^{1/2}(\hat{\beta}_0 - \beta)$ é assintoticamente normal multivariado com média 0 e matriz de covariância $V_0 = I_0^{-1}I_0^*I_0^{-1}$, de modo que I_0 é a matriz de informação de Fisher e $I_0^* = \mathbb{E}(SS')$.

Em contrapartida, o estimador $\hat{\beta}_1$ fundamenta-se na hipótese de que cada série binária constitui a realização de uma cadeia de Markov estacionária pelas quais

$$\text{logit } (\pi_i) = Z'_i\beta \quad (2.1)$$

e

$$\rho = \text{Corr}(Y_{it}, Y_{i,t-1} | Z_i). \quad (2.2)$$

Uma importante implicação de (2.2) diz respeito à hipótese assumida pelos autores, a saber, de que as observações Y_{it} , admitem uma mesma medida de autocorrelação de ordem 1. Nessa linha de raciocínio, Zeger et al. (1985) estabelecem a esperança (condicional) de Y_{it} também como função de uma medida de correlação, dada por

$$\pi_{it} = \mathbb{E}(Y_{it}|Y_{i,t-1}, Z_i) = \pi_i + \rho(Y_{i,t-1} - \pi_i). \quad (2.3)$$

Algumas considerações podem ser feitas acerca do termo de (2.3), que passa a fazer sentido somente a partir da segunda observação (binária). Nesse caso, π_{i2} , por exemplo, constituiria um valor esperado de Y_{i2} , $\mathbb{E}(Y_{i2}|Y_{i,1}, Z_i)$, o qual também pode ser pensado como a probabilidade de $\mathbb{P}(Y_{i,2} = 1|Z_i)$ mais uma medida de correlação da diferença entre o resultado anterior ($Y_{i,1}$) e a $\mathbb{P}(Y_{i,2} = 1|Z_i)$.

Assim, e de forma intencional, a função log-verossimilhança é apresentada explicitamente como uma composição entre a primeira observação, $y_{i1} \log(\pi_i) + (1 - y_{i1}) \log(1 - \pi_i)$, e o somatório das demais observações, $\sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\}$, passando por todas as K UEs. Em outros termos,

$$\begin{aligned} \ell_1(\beta, \rho) = \sum_{i=1}^K & \left[y_{i1} \log \pi_i \right. \\ & + (1 - y_{i1}) \log(1 - \pi_i) \\ & \left. + \sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\} \right]. \end{aligned} \quad (2.4)$$

Com relação especificamente aos termos da primeira observação, é possível notar que

$$\begin{aligned} & y_{i1} \log(\pi_i) + (1 - y_{i1}) \log(1 - \pi_i) \\ & = y_{i1} \log(\pi_i) + \log(1 - \pi_i) - y_{i1} \log(1 - \pi_i) \end{aligned} \quad (2.5)$$

$$\begin{aligned}
&= y_{i1}[\log(\pi_i) - \log(1 - \pi_i)] + \log(1 - \pi_i) \\
&= y_{i1} \left[\log \left(\frac{\pi_i}{1 - \pi_i} \right) \right] + \log(1 - \pi_i) \\
&= y_{i1}[\text{logit}(\pi_i)] + \log(1 - \pi_i) \\
&= y_{i1}Z_i'\beta + \log(1 - \pi_i).
\end{aligned}$$

Uma vez que $\pi_i = \frac{e^{Z_i\beta}}{1+e^{Z_i\beta}}$, temos que $\log(1 - \pi_i) = \log \left(1 - \frac{e^{Z_i\beta}}{1+e^{Z_i\beta}} \right) = \log \left(\frac{1}{1+e^{Z_i\beta}} \right) = -\log(1 + e^{Z_i\beta})$, de modo que

$$y_{i1}Z_i'\beta + \log(1 - \pi_i) = y_{i1}Z_i'\beta - \log(1 + e^{Z_i\beta}), \quad (2.6)$$

fazendo com que a função log-verossimilhança (2.4) assuma a forma

$$\begin{aligned}
\ell_1(\beta, \rho) = \sum_{i=1}^K & \left[y_{i1}Z_i'\beta \right. \\
& - \log(1 + e^{Z_i\beta}) \\
& \left. + \sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\} \right], \quad (2.7)
\end{aligned}$$

tal qual em Zeger et al. (1985).

As equações de escore para β e ρ passam, então, a assumir a forma, para $u = 1, \dots, s$,

$$\begin{aligned}
0 = S_u(\beta, \rho) &= \frac{\partial \ell_1}{\partial \beta_u} \\
&= \sum_{i=1}^K \left\{ Z_{iu}(y_{i1} - \pi_i) \right. \\
&\quad \left. + (1 - \rho) \sum_{t=2}^{n_i} \frac{(Z_{iu}\pi_i\bar{\pi}_i)(y_{it} - \pi_{it})}{\pi_{it}\bar{\pi}_{it}} \right\}, \quad (2.8)
\end{aligned}$$

sendo $\bar{\pi}_i = 1 - \pi_i$ e $\bar{\pi}_{it} = 1 - \pi_{it}$, para o caso de β e

$$0 = S_{s+1}(\beta, \rho) = \frac{\partial \ell_1}{\partial \rho} = \sum_{i=1}^K \sum_{t=2}^{n_i} \frac{(y_{it} - \pi_{it})(y_{i,t-1} - \pi_{i,t-1})}{\pi_{it}\bar{\pi}_{it}}, \quad (2.9)$$

para o caso de ρ . As demonstrações de (2.8) e (2.9) encontram-se no apêndice deste capítulo.

A matriz de informação de Fisher, que assumiremos I_1 , será dada pela matriz particionada

$$I_1 = \begin{bmatrix} I_{\beta\beta} & I_{\beta\rho} \\ I_{\rho\beta} & I_{\rho\rho} \end{bmatrix},$$

constituída fundamentalmente por elementos de derivadas de segunda ordem,

seja a partir de $S_u(\beta, \rho)$, seja a partir de $S_{s+1}(\beta, \rho)$. Diante desses resultados, os autores Zeger et al. (1985) apresentaram uma segunda proposição:

Proposição 2. Seja Y_{it} ($t = 1, \dots, n_i < \infty; i = 1, \dots, K$) uma série binária estacionária tal que $\text{logit } \mathbb{E}(Y_{it}|Z_i) = Z_i'\beta$ e $\text{Corr}(Y_{it}, Y_{i,t-1}) = \rho$. Então, $K^{1/2}(\hat{\beta}_1 - \beta)$ é assintoticamente

normal multivariado, quando $K \rightarrow \infty$, com média 0 e matriz de covariância $V_1 = I_1^{-1} I_1^* I_1^{-1}$, de modo que I_1 (descrita anteriormente) é a matriz de informação de Fisher e $I_1^* = E(SS')$.

É importante notar que as proposições 1 e 2, necessárias tanto para as estimativas quanto para a inferência de $\hat{\beta}_0$ e $\hat{\beta}_1$, as matrizes da informação de Fisher I_0^* e I_1^* , possuem, ambas, a mesma definição, isto é, são ambas iguais a $E(SS')$, ou seja, iguais ao valor esperado do quadrado da derivada do logaritmo da função de verossimilhança (ainda que, sob certas condições, também possa ser escrita como o valor esperado da derivada segunda do log-verossimilhança).

Em razão do exposto, a esperança $E(SS')$, necessária para a obtenção de I_0^* e I_1^* , e por tabela, também de $V_0 = I_0^{-1} I_0^* I_0^{-1}$ e de $V_1 = I_1^{-1} I_1^* I_1^{-1}$, exige o conhecimento do verdadeiro modelo (log-verossimilhança) para o conjunto de dados. Esse fato também permite afirmar, por meio de (2.8) e (2.9), que o conhecimento da verdadeira dependência temporal (correlação) também é necessário, uma vez que a medida de correlação está compreendida na própria função log-verossimilhança. Dessa forma, dificuldades técnicas surgem no que diz respeito ao processo de estimação de I_0^* e I_1^* .

Zeger et al. (1985) sugerem fazer uso das informações das replicações das próprias UEs no intuito de estimar I_0^* e I_1^* com base nas proposições 1 e 2. Dessa forma, as estimativas das variâncias para β_0 e β_1 poderiam ser obtidas por meio do produto cruzado de funções de verossimilhança aproximadas (de n -ésima ordem), utilizadas, segundo os autores, em Azzalini (1983) e Ogata (1980).

De acordo com Zeger et al. (1985), a verdadeira função log-verossimilhança para a i -ésima UE, $\ell_i(\theta)$, é aproximada por

$$\ell_i^{(n)} = \sum_{t=1}^{n_i} \log f(y_{it} | y_{i,t-1}, \dots, y_{i,t-n}; \theta), \quad (2.10)$$

com $\ell_i^{(n)}$ a função log-verossimilhança de n -ésima ordem. Nesse caso, de acordo ainda com os autores, a consistência de um estimador $\hat{\beta}$ genérico exige que a esperança condicional $E(Y_{it} | Y_{i,t-1}, \dots, Y_{i,t-n}; Z_i)$ contenha, para as distintas UEs, $2^n - 1$ parâmetros em comum, diferindo apenas com relação a π_i .

Dessa maneira, I^* poderia ser estimado por

$$\hat{I}_{uv}^* = \frac{1}{K} \sum_{i=1}^K \left(\frac{\partial \ell_i}{\partial \beta_u} \frac{\partial \ell_i}{\partial \beta_v} \right), \quad (2.11)$$

independentemente da estrutura de dependência temporal, apesar de que, segundo os autores, a exigência da esperança condicional envolvida em ℓ_i seja elevada e, em razão disso, aproximações de ordens mais baixas sejam preferíveis.

Por fim, em suas discussões finais, Zeger et al. (1985) chamam ainda a atenção para o fato de que estimativas inconsistentes podem ser obtidas quando séries forem indevidamente caracterizadas como cadeias de Markov de ordem p quando, na realidade, a ordem (do processo) é superior.

2.4 O modelo GEE com função de verossimilhança

No artigo “Longitudinal data analysis using generalized linear models”, Liang e Zeger (1986) desenvolvem uma nova classe de modelos longitudinais, GEE. Por meio de uma nova metodologia de estimação, os GEE permitem não somente um tipo de resposta (binária), mas uma ampla gama de possibilidades (modelos) para a variável resposta (família exponencial) tais como: Bernoulli, binomial, Poisson, normal, gama, beta, multinomial etc.

De certa maneira, é possível afirmar que os autores estenderam a classe dos modelos lineares generalizados (MLG) para a análise de dados longitudinais.

Os modelos MLG foram inicialmente propostos no valoroso artigo de Nelder e Wedderburn (1972), mas uma descrição mais detalhada de suas propriedades e características pode ser encontrada no livro de McCulloch e Nelder (1989).

A classe dos modelos GLM possui, como premissa fundamental, a independência entre as observações. Dessa maneira, uma grande inovação, por assim dizer, da classe dos modelos GEE consiste na possibilidade de se levar em consideração a correlação (possivelmente existente) entre as observações de uma mesma UE ao longo do tempo.

Diferentemente de Zeger et al. (1985), Liang e Zeger (1986) introduzem uma matriz de correlação nas equações de estimação comumente empregadas no processo de estimação de GLM, realizando uma modificação/adaptação em sua forma/expressão analítica. Segundo Fitzmaurice, Laird e Ware (2011): “*The essential idea behind the GEE approach is to generalize and extend the usual likelihood equations for a generalized linear model for a univariate response by incorporating the covariance matrix of the vector of responses, Y_i .*” (p. 354)

A seguir serão detalhados os procedimentos adotados por Liang e Zeger (1986) com a finalidade de compreender melhor as inovações trazidas pelos modelos GEE.

Na Seção 1, os autores apresentam algumas notações, tais como a variável resposta $Y_i = (y_{i1}, \dots, y_{in_i})^T$ constituindo um vetor $n_i \times 1$ e $X_i = (x_{i1}, \dots, x_{in_i})^T$ uma matriz de covariáveis de ordem $n_i \times p$ para o i -ésimo indivíduo ($i = 1, \dots, K$).

Fitzmaurice et al. (2011, p. 343) representam precisamente essas quantidades na seguinte forma matricial:

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix}, i = 1, \dots, K \quad (2.12)$$

E

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in_i} \end{pmatrix} = \begin{pmatrix} x_{i11} & x_{i12} & \dots & x_{i1p} \\ x_{i21} & x_{i22} & \dots & x_{i2p} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ x_{in_i1} & x_{in_i2} & \Lambda & x_{in_ip} \end{pmatrix}, i = 1, \dots, K, \quad (2.13)$$

sendo que $x_{i1}, x_{i2}, \dots, x_{in_i}$ são vetores, cada qual expressando p medições (ou p diferentes covariáveis) em cada ponto do tempo (genericamente em n_i pontos do tempo).

A partir dessas representações matriciais, é possível notar a importância que os autores atribuem para o fato de que o interesse científico reside no padrão de mudança ao longo do tempo (no caso, a dependência da variável resposta em relação às covariáveis).

Fitzmaurice et al. (2011, p. 343) atentam também para o fato de que a matriz X_i contém covariáveis que podem variar (ou não) ao longo do tempo. No caso das primeiras, as covariáveis são *intraindivíduos* (within-subject), enquanto que aquelas que não variam ao longo do tempo seriam as *entre-indivíduos* (between-subject).

Fitzmaurice et al. (2011, p. 264) ilustram com um exemplo aplicado a importância de se levar em consideração uma medida de correlação em um contexto de dados longitudinais. Os autores sugerem uma comparação entre dois planejamentos experimentais: i) um estudo longitudinal cujo interesse consiste em medir mudanças em uma variável resposta em um esquema do tipo “antes” e “depois” e ii) um estudo transversal (*cross-section*) no qual participantes são aleatoriamente alocados em um grupo (controle ou tratamento).

Em ambos os esquemas experimentais, são duas as notas passíveis de serem obtidas de ambas as observações (Y_{i1} e Y_{i2}). A variabilidade das notas oriundas do estudo longitudinal é dada por

$$Var(Y_{i2} - Y_{i1}) = Var(Y_{i1}) + Var(Y_{i2}) - 2Cov(Y_{i1}, Y_{i2}) \quad (2.14)$$

$$\begin{aligned}
&= \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \\
&= \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2,
\end{aligned}$$

uma vez que $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ e a covariância pode ser escrita como $Cov(X,Y) = \rho_{X,Y}\sigma_X\sigma_Y$.

Já no caso do estudo transversal, a variabilidade das notas seria dada por

$$\begin{aligned}
Var(Y_{i2} - Y_{i1}) &= Var(Y_{i1}) + Var(Y_{i2}) \\
&= \sigma_1^2 + \sigma_2^2,
\end{aligned} \tag{2.15}$$

sendo que, nesse caso, as variâncias dizem respeito a indivíduos distintos pertencentes a grupos distintos.

No primeiro caso, a variância é *intra-indivíduo*, enquanto que, no segundo esquema, é *entre-indivíduos*. Supondo que a correlação entre as medidas repetidas seja positiva, a variância no estudo longitudinal é sempre menor quando comparada com a variância do estudo transversal.

Fitzmaurice et al. (2011, p. 265) fazem, ainda, uma outra suposição, a saber, que as variâncias sejam iguais em ambos os estudos, isto é, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Isso faria com que a variância no estudo longitudinal fosse de $2\sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1 - \rho)$ e, no caso do estudo transversal, fosse de $2\sigma^2$. Caso $\rho = 0$, as variâncias em ambos os estudos seriam as mesmas. Mas, em um cenário mais provável, se $0 < \rho < 1$, a variância no estudo longitudinal seria sempre menor, uma fração da variância relativa ao estudo transversal, refletindo uma maior precisão por parte do estudo longitudinal.

Especificamente com relação ao número de observações de cada UE, Fitzmaurice et al. (2011) observam que, se $n_i = 1$, ou seja, se em cada UE existisse uma única observação, então um GLM poderia naturalmente ser aplicado. Mas com medidas repetidas em uma mesma UE, as correlações entre as observações devem receber atenção especial.

No artigo, Liang e Zeger (1986) iniciam caracterizando a variável resposta como pertencendo à família exponencial, isto é,

$$f(y_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi], \tag{2.16}$$

com θ_{it} sendo um parâmetro desconhecido e função de um preditor linear, $\theta_{it} = h(\eta_{it})$, $\eta_{it} = x_{it}\beta$ e um parâmetro de escala, ϕ , podendo, eventualmente, $\phi = 1$ e duas propriedades bastante conhecidas acerca da família exponencial, a saber, a esperança e a variância (no caso, da variável resposta) descritas como primeira e segunda derivadas, respectivamente, da função densidade de probabilidade.

$$E(y_{it}) = a'(\theta_{it}) \tag{2.17}$$

e

$$Var(y_{it}) = \frac{a''(\theta_{it})}{\phi}. \quad (2.18)$$

Na Seção 2, os autores apresentam um estimador $\hat{\beta}_I$ para β , sob a suposição de que as observações repetidas de uma mesma UE são independentes. Logo, supondo independência entre as K UEs, a solução para $\hat{\beta}_I$ consistiria em uma soma de K equações score, semelhante ao caso dos GLM que, nesse caso, a soma dar-se-ia pelo somatório de todas as n observações (independentes), ou seja,

$$U_I(\beta) = \sum_{i=1}^K X_i^T \Delta_i S_i = 0, \quad (2.19)$$

sendo Δ_i uma matriz diagonal com elementos $\frac{d\theta_{it}}{d\eta_{it}}$ e $S_i = Y_i - a'(\theta)$.

A exposição da semelhança da solução encontrada por Liang e Zeger em (2.19) com o método empregado nos modelos GLM encontra-se no apêndice desse capítulo (seção 2.6).

Em seguida, Liang e Zeger (1986) definem uma matriz diagonal com informações de variância de cada UE (derivada segunda de um dos termos da função densidade na forma exponencial, no nível da UE) e apresentam um teorema a partir do qual fica comprovada sua consistência (assintótica) de $\hat{\beta}_I$, quando K cresce indefinidamente, com média nula e matriz de covariância, V_I , escrita como função de

$$V_I = \lim_{K \rightarrow \infty} K \left(\sum_{i=1}^K X_i^T \Delta_i A_i \Delta_i X_i \right)^{-1} \left(\sum_{i=1}^K X_i^T \Delta_i Cov(Y_i) \Delta_i X_i \right) \left(\sum_{i=1}^K X_i^T \Delta_i A_i \Delta_i X_i \right)^{-1}. \quad (2.20)$$

Por meio de (2.20), é possível perceber que não é necessário conhecer ϕ para estimar V_I (ainda que essa seja uma função daquela).

Os autores chamam a atenção para o fato de que o estimador $\hat{\beta}_I$ e a $Var(\hat{\beta}_I)$ são consistentes quando fornecida uma correta especificação para a regressão. Como vantagem, $\hat{\beta}_I$ pode ser facilmente estimado (com software adequado) e, como desvantagem, $\hat{\beta}_I$ pode apresentar problemas de eficiência na presença de autocorrelação.

A Seção 3 subdivide-se em três partes: Seção 3.1 (geral), Seção 3.2 (conexão com o método Gauss-Newton) e Seção 3.3 (estimadores de α e ϕ).

Na Seção 3.1, os autores introduzem um conjunto de equações de estimação “generalizadas”, isto é, que levam em consideração uma matriz de correlação (na realidade, uma matriz de covariância) no intuito de melhor lidar com o problema da eficiência. Nessa seção, o objetivo foi desenvolver uma nova classe de estimadores, $\hat{\beta}_G$, com variâncias passíveis de serem estimadas, V_G .

Liang e Zeger (1986) passaram a considerar uma matriz de variância que expressasse as informações de cada UE e que também fosse função de uma matriz de correlação “de trabalho” $R(\alpha)$ dessa mesma UE, dada por

$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi, \quad (2.21)$$

com α sendo um vetor $s \times 1$, de modo a representar uma matriz de correlação típica. De acordo com Liang e Zeger (1986), se a matriz $R(\alpha)$ coincidir com a verdadeira matriz de correlação dos Y_i 's, a matriz V_i passa a coincidir com a matriz $Cov(Y_i)$.

No caso da expressão (2.21), o índice i expressa a possibilidade de diferentes dimensões da matriz V_i entre as UEs.

Repare que, nesse caso, a matriz de variância V_i passa a depender da matriz de correlação $R(\alpha)$ e as equações de estimação “generalizadas” também passam a expressar esse fato, de modo que a solução para $\hat{\beta}_G$ é dada por

$$\sum_{i=1}^K D_i^T V_i^{-1} S_i = 0. \quad (2.22)$$

com $D_i = \frac{d\{a_i'(\theta)\}}{d\beta} = A_i \Delta_i X_i$.

Uma análise da semelhança da solução encontrada por Liang e Zeger em (2.22) com o método comumente empregado nos modelos GLM também se encontra na seção 2.6.

Não obstante, os autores chamam a atenção para dois fatos importantes:

- a equação (2.22), solução de $\hat{\beta}_G$, reduz-se à equação (2.19), solução de $\hat{\beta}_I$, se a matriz $R(\alpha)$ reduzir-se à matriz identidade.
- para cada i , $U_i(\beta, \alpha) = D_i^T V_i^{-1} S_i$ se assemelha à solução obtida do método proposto pelos criadores dos GLM, $U(\beta)$, exceto pelo fato de que $U_i(\beta, \alpha)$ depende, também, da autocorrelação α , via matriz V_i .

Em seguida, Liang e Zeger (1986) comprovaram uma propriedade amostral (propriedade das grandes amostras) de $\hat{\beta}_G$ por meio do teorema 2 no artigo publicado por eles. Com base em algumas suposições (dentre elas a consistência), a distribuição de $K^{1/2}(\hat{\beta}_G - \beta)$ é assintoticamente normal multivariada, com média zero e matriz de covariância dada por

$$V_G = \lim_{K \rightarrow \infty} K \left(\sum_{i=1}^K D_i^T V_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^K D_i^T V_i^{-1} Cov(Y_i) V_i^{-1} D_i \right\} \left(\sum_{i=1}^K D_i^T V_i^{-1} D_i \right)^{-1}. \quad (2.23)$$

Liang e Zeger (1986) chamam, ainda, a atenção para alguns fatos:

- a consistência de $\hat{\beta}_G$ e V_G depende apenas da correta especificação da média e não da correta escolha de R ;
- a variância assintótica de $\hat{\beta}_G$ também não depende da escolha do estimador para α e ϕ (dentre aqueles $K^{1/2}$ –consistentes).

Jang (2011, p. 3), por sua vez, reescreve (2.23) como

$$V_G = \lim_{K \rightarrow \infty} K M_0^{-1} M_1 M_0^{-1} \quad (2.24)$$

e analisa que Liang e Zeger propõem que, se $Cov(\hat{\beta})$ for estimada por meio do estimador “sanduíche”

$$\hat{\Sigma}_S = \hat{M}_0^{-1} \hat{M}_1 \hat{M}_0^{-1}, \quad (2.25)$$

sendo $\hat{M}_0 = \sum_{i=1}^K \hat{D}_i^T V_i(\hat{\alpha}, \hat{\phi})^{-1} \hat{D}_i$, $\hat{M}_1 = \left[\sum_{i=1}^K \hat{D}_i^T V_i(\hat{\alpha}, \hat{\phi})^{-1} (Y_i - \mu_i(\hat{\beta})) (Y_i - \mu_i(\hat{\beta}))^T V_i(\hat{\alpha}, \hat{\phi})^{-1} \hat{D}_i \right]$, $\hat{\alpha}$ e $\hat{\phi}$ estimadores $\sqrt{n}$ -consistentes para α e ϕ , então, a diferença entre $K\hat{\Sigma}_S$ e V_G converge para zero, em probabilidade.

Com relação aos valores faltantes (*missing values*), tanto no caso de $\hat{\beta}_I$ e $\hat{V}_I$ quanto no caso de $\hat{\beta}_G$ e $\hat{V}_G$, é esperado que se processem (caso ocorram) em um padrão completamente ao acaso (MCAR).

Na Seção 3.2 (conexão com o método Gauss-Newton), os autores sugeriram um método iterativo adaptado para estimadores de equações generalizadas. Pelo fato de a classe dos modelos GEE depender tanto de β quanto dos parâmetros α e ϕ e ainda não possuir uma solução com forma fechada, métodos iterativos passam a ser necessários (Jang, 2011, p. 3).

O método descrito em Liang e Zeger (1986) possui uma pequena modificação no método de *Score de Fisher* para um vetor de parâmetros β , e estimação pelo método dos momentos para α e ϕ . Desse modo, temos que

$$\begin{aligned} & \hat{\beta}_{j+1} \\ &= \hat{\beta}_j - \left\{ \sum_{i=1}^K D_i^T(\hat{\beta}_j) \tilde{V}_i^{-1}(\hat{\beta}_j) D_i(\hat{\beta}_j) \right\}^{-1} \left\{ \sum_{i=1}^K D_i^T(\hat{\beta}_j) \tilde{V}_i^{-1}(\hat{\beta}_j) S_i(\hat{\beta}_j) \right\}, \end{aligned} \quad (2.26)$$

sendo que as estimativas obtidas na j -ésima iteração servirão de base para as estimativas da próxima iteração (método *iteratively reweighted least squares* – IRLS) até que um determinado critério de convergência seja satisfeito.

Na Seção 3.3 (estimadores de α e ϕ), os autores propõem estimadores tanto para α quanto para ϕ , os quais podem ser obtidos, em uma dada iteração, a partir dos resíduos de Pearson estimados, definidos por

$$\hat{r}_{it} = \frac{y_{it} - a'(\hat{\theta}_{it})}{a''(\hat{\theta}_{it})^{1/2}}, \quad (2.27)$$

sendo que $\hat{\theta}_{it}$ depende do valor atual de β . Para o caso do parâmetro de dispersão ϕ ,

$$\hat{\phi}^{-1} = \sum_{i=1}^K \sum_{t=1}^{n_i} \frac{\hat{r}_{it}^2}{N - p}, \quad (2.28)$$

de modo que $N = \sum n_i$. Para uma estimativa consistente de α , é proposto

$$\hat{R}_{uv} = \sum_{i=1}^K \frac{\hat{r}_{iu}\hat{r}_{iv}}{N - p}. \quad (2.29)$$

2.5 O modelo GEE com função de quase-verossimilhança

Nesse novo artigo, Zeger e Liang (1986) inovaram ao permitir que os estimadores da classe dos modelos GEE também pudessem ser obtidos por meio da função de quase-verossimilhança (e não a função de verossimilhança). Apesar de se tratar dos mesmos autores do tema principal da seção anterior (Seção 2.4) (Liang & Zeger, 1986), o assunto difere consideravelmente.

Da perspectiva do usuário da técnica, a utilidade em estimar modelos GEE via função de quase-verossimilhança reside, fundamentalmente, no fato de que não mais é necessário ter o conhecimento específico acerca da função de verossimilhança a ser utilizada para a variável resposta do estudo. A escolha por utilizar a função de quase-verossimilhança dispensa o conhecimento da distribuição da variável resposta.

Para utilizar a função de quase-verossimilhança, em um contexto de estimação de modelos GEE, é necessário o estabelecimento de uma determinada função que seja capaz de relacionar a média com a variância dos dados.

Zeger e Liang (1986) atribuem a McCullagh e Nelder (1983), a observação de que: *“This partial specification of the outcome distribution leads to simple techniques for regression analyses of Gaussian, gamma, Poisson, binomial, categorical, and ordinal data.”*

Com relação à possibilidade de a variável resposta não possuir distribuição normal, existiam poucas técnicas, até então. Modelos de efeitos aleatórios com resposta binária eram a exceção. Exemplos citados são os modelos de efeitos aleatórios para observações seriais com resposta binária de Stiratelli, Laird e Ware (1984), modelos de componentes de variância com resposta binária de Anderson e Aitkin (1985), modelos *probit* de Oichi e Prentice (1984) e modelos de medidas repetidas de dados categóricos de Koch et al. (1971), dos quais, apenas os modelos logísticos de efeitos aleatórios de Stiratelli et al. (1984) e de Anderson e Aitkin (1985).

Em seguida, Zeger e Liang (1986) aludem a um artigo de Cox (1970) a partir do qual se introduz um modelo de regressão logístico para a esperança condicional de dados binários, dadas as observações anteriores. Fundamentava-se, a realidade, em uma cadeia de Markov na qual as probabilidades de transição dependiam das covariáveis.

Vale lembrar que o próprio modelo desenvolvido por Zeger et al. (1985) constituía uma classe de modelos de Markov para dados binários longitudinais, mas, possuía uma restrição, a saber, apesar dos coeficientes de regressão (bem como suas variâncias) poderem ser obtidos sob certas suposições já analisadas na Seção 2.3.

Em seguida, conforme já descrito, Liang e Zeger (1986) combinam características da classe de modelos desenvolvidos em Zeger et al. (1985) com a estrutura da classe de modelos GLM, explicitando que, no caso, o objetivo principal consiste em propor que a estimação de modelos GEE pudesse ser realizada por meio das funções de quase-verossimilhança (e não por meio das funções de verossimilhança).

Nas Seções 2 e 3 do artigo de Zeger e Liang (1986), os autores introduziram a lógica da função de quase-verossimilhança (Seção 2) e a conectaram com os procedimentos de estimação de modelos GEE (Seção 3).

Na Seção 2, Zeger e Liang (1986) apontaram que Wedderburn (1974) inicialmente propôs o conceito de quase-verossimilhança, o qual exigiria poucas suposições sobre a distribuição da variável dependente em questão para o emprego dessa metodologia em modelos de regressão. Não obstante, permitiria, ainda, ser utilizada em uma multiplicidade de casos.

Wedderburn (1974) estabelece que:

To define a likelihood we have to specify the form of distribution of the observations, but to define a quasi-likelihood function, we need only specify a relation between the mean and the variance of the observations and the quasi-likelihood can be used for estimation. (p. 439)

Logo em seguida, Wedderburn (1974) define a função de quase-verossimilhança como uma função $K(y_i, \mu_i)$ tal que

$$\frac{\partial K(y_i, \mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{V(\mu_i)}. \quad (2.30)$$

ou, de forma equivalente,

$$K(y_i, \mu_i) = \int^{\mu_i} \frac{y_i - \mu'_i}{V(\mu'_i)} d\mu'_i + \text{função de } y_i. \quad (2.31)$$

Além disso, apresenta (com demonstrações) quatro propriedades:

1. $\mathbb{E} \left(\frac{\partial K}{\partial \mu} \right) = 0;$

2. $\mathbb{E}\left(\frac{\partial K}{\partial \beta_i}\right) = 0;$
3. $\mathbb{E}\left(\frac{\partial K}{\partial \mu}\right)^2 = -\mathbb{E}\left(\frac{\partial^2 K}{\partial \mu^2}\right) = \frac{1}{V(\mu)};$
4. $\mathbb{E}\left(\frac{\partial K}{\partial \beta_i} \frac{\partial K}{\partial \beta_j}\right) = -\mathbb{E}\left(\frac{\partial^2 K}{\partial \beta_i \partial \beta_j}\right) = \frac{1}{V(\mu)} \frac{\partial \mu}{\partial \beta_i} \frac{\partial \mu}{\partial \beta_j}.$

McCullagh e Nelder (1989) destinam um capítulo inteiro (Capítulo 9) à função de quase-verossimilhança, $Q(\mu, y)$, e a apresenta genericamente como uma função

$$Q(\mu, y) = \int_y^\mu \frac{y - t}{\sigma^2 V(t)} dt \quad (2.32)$$

e que, ao admitir independência entre as observações, a quase-verossimilhança para a totalidade dos dados seria dada por

$$Q(\boldsymbol{\mu}, \mathbf{y}) = \sum_i Q_i(\mu_i, y_i). \quad (2.33)$$

Em seguida, McCullagh e Nelder (1989) introduzem a função de *quase-desvio* (*quasi-deviance*), frequentemente utilizada para estimação de parâmetros em um contexto de função de quase-verossimilhança. Em essência, a função quase-desvio diz respeito a uma transformação da própria função de quase-verossimilhança. A função desvio referente a uma única observação é dada por

$$D(y; \mu) = -2\sigma^2 Q(\mu_i, y_i) = 2 \int_\mu^y \frac{y - t}{V(t)} dt. \quad (2.34)$$

Na sequência, McCullagh e Nelder (1989, p. 326) apresentam uma tabela (Tabela 9.1) com algumas funções de quase-verossimilhança que, dependendo da expressão de $V(\mu)$, exibiram comportamentos característicos de algumas distribuições conhecidas. Vale lembrar que, segundo Zeger e Liang (1986), “*Many, but not all, of these quasi-likelihood correspond to real log likelihood for known distributions.*” (p. 325)

Retomando Zeger e Liang (1986), precisamente no final da Seção 2, os autores apresentam o estimador de quase-verossimilhança como uma solução de um sistema de equações de *quase-score*.

É importante enfatizar que uma função score refere-se à derivada do logaritmo de uma função densidade de probabilidade. Nessa linha de raciocínio, *quase-score* pode ser entendido como a derivada de uma função de quase-verossimilhança, uma em particular que relacione a média μ com β , como, por exemplo, $\mu(\beta)$. Consideremos, de acordo com a definição de Wedderburn (1974),

$$\frac{\partial Q_i(\mu_i)}{\partial \beta_r} = \frac{\partial Q_i(\mu_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_r} = \frac{y_i - \mu_i}{\sigma^2 V_i(\mu_i)} D_{ir}, \quad (2.35)$$

de modo que $D_{ir} = \frac{\partial \mu_i}{\partial \beta_r}$. Logo, ao levar consideração todas as observações, teremos

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \beta_r} = \sum_{i=1}^n \frac{\partial Q_i(\mu_i)}{\partial \beta_r} = \sum_{i=1}^n \frac{y_i - \mu_i}{\sigma^2 V_i(\mu_i)} D_{ir}. \quad (2.36)$$

Do ponto de vista matricial, temos que

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \beta_r} = \frac{1}{\sigma^2} [D_{1r} \quad D_{2r} \quad \dots \quad D_{nr}] \begin{bmatrix} V_1^{-1}(\mu_1) & 0 & \dots & 0 \\ 0 & V_2^{-1}(\mu_2) & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & V_n^{-1}(\mu_n) \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ \vdots \\ y_n - \mu_n \end{bmatrix} \quad (2.37)$$

ou, de modo mais resumido,

$$\frac{\partial Q(\boldsymbol{\mu})}{\partial \beta_r} = \frac{1}{\sigma^2} D_r^T V^{-1}(\mathbf{y} - \boldsymbol{\mu}), \quad (2.38)$$

sendo $D_r = [D_{1r} \quad D_{2r} \quad \Lambda \quad D_{nr}]^T$. Dessa maneira, podemos escrever que a função quase-score $U(\boldsymbol{\beta})$, foi obtida através da diferenciação de $\frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\beta}}$ e que é solução para $\boldsymbol{\beta}$. Em outros termos,

$$U(\boldsymbol{\beta}) = \frac{\partial Q(\boldsymbol{\mu})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} D^T V^{-1}(\mathbf{y} - \boldsymbol{\mu}), \quad (2.39)$$

tal que $U(\hat{\boldsymbol{\beta}}) = 0$. A matriz D , de ordem $n \times p$, é dada por

$$D_{n \times p} = [D_1 \quad D_2 \quad \Lambda \quad D_p] = \begin{bmatrix} D_{11} & D_{12} & \Lambda & D_{1p} \\ D_{21} & D_{22} & \Lambda & D_{2p} \\ \text{M} & \text{M} & \text{O} & \text{M} \\ D_{n1} & D_{n2} & \Lambda & D_{np} \end{bmatrix}. \quad (2.40)$$

A matriz de covariância de $U(\boldsymbol{\beta})$, de acordo com McCullagh e Nelder (1989), condiz com o valor esperado negativo de $\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, que é igual a $D^T V^{-1} D / \sigma^2$, resultado esse que pode ser obtido por

$$\begin{aligned} \text{Cov}[U(\boldsymbol{\beta})] &= \text{Var}[U(\boldsymbol{\beta})] \\ &= \mathbb{E}[U(\boldsymbol{\beta}) U^T(\boldsymbol{\beta})] \\ &= -\mathbb{E} \left[\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right] \\ &= D^T V^{-1} D / \sigma^2, \end{aligned} \quad (2.41)$$

uma vez que, sob certas condições apontadas por McCullagh e Nelder (1989, pp. 324-325), as propriedades do logaritmo da função de verossimilhança também se aplicam para o caso das

funções de quase-verossimilhança. Como a informação de Fisher é, por definição, a variância da função score, temos que, nesse caso, a matriz $D^T V^{-1} D / \sigma^2$ exerce papel equivalente à informação de Fisher, observação feita pelos autores, os quais definem

$$\mathbf{i}_\beta = D^T V^{-1} D / \sigma^2. \quad (2.42)$$

Sob certas condições de limites usuais dos autovalores de $\mathbf{i}_\beta$, a matriz assintótica de covariância de $\hat{\beta}$ será dada por

$$\text{Cov}(\hat{\beta}) \approx \mathbf{i}_\beta^{-1} = \sigma^2 (D^T V^{-1} D)^{-1}. \quad (2.43)$$

Desse modo, pelo método Newton-Raphson, e supondo um valor inicial $\hat{\beta}_0$ suficientemente próximo de $\hat{\beta}$, teremos

$$\hat{\beta}_1 = \hat{\beta}_0 + \sigma^2 (\hat{D}_0^T \hat{V}_0^{-1} \hat{D}_0)^{-1} (\sigma^2)^{-1} \hat{D}_0^T \hat{V}_0^{-1} (\mathbf{y} - \hat{\mu}_0). \quad (2.44)$$

Ao cancelarmos σ^2 , tem-se

$$\hat{\beta}_1 = \hat{\beta}_0 + (\hat{D}_0^T \hat{V}_0^{-1} \hat{D}_0)^{-1} \hat{D}_0^T \hat{V}_0^{-1} (\mathbf{y} - \hat{\mu}_0), \quad (2.45)$$

de modo que a estimativa de quase-verossimilhança $\hat{\beta}$ será obtida iterativamente, segundo um critério de convergência.

Zeger e Liang (1986) finalizam a Seção 2 exibindo o estimador de *quase-verossimilhança* como solução de um sistema de equações *quase-score* dado por

$$S_k = \sum_{i=1}^K \frac{\partial \mu_i}{\partial \beta_k} v_i^{-1} (y_i - \mu_i) = 0, \quad k = 1, \dots, p, \quad (2.46)$$

tal que $v_i = g(\mu_i) / \phi$, sendo g uma função conhecida, dependente da média μ_i . Além disso, os autores chamam a atenção para o fato de que S_k representa equações de *score* quando a variável resposta y_i possui distribuição pertencente à família exponencial.

No caso, é possível perceber a equivalência de S_k com a expressão obtida mais acima, $\frac{1}{\sigma^2} D_r^T V^{-1} (\mathbf{y} - \boldsymbol{\mu})$, com a diferença de que S_k lida com várias observações de um mesmo indivíduo e a solução de quase-verossimilhança envolve todas n observações independentes, ainda em um contexto de GLM.

Na Seção 3 de Zeger e Liang (1986), foi realizada uma adaptação da função de *quase-verossimilhança* para uma estimação adequada de modelos GEE, com dados longitudinais.

De certa forma, a Seção 3 reproduz os mesmos passos adotados no processo de estimação do artigo anterior (Liang e Zeger, 1986), com a diferença de que, nesse caso, o método é adaptado para as funções de *quase-verossimilhança*.

Como foi possível notar na Seção 2.4 deste capítulo, Liang e Zeger (1986) introduziram uma matriz de correlação $R(\alpha)$ e adaptaram o método de estimação no intuito de obter estimativas.

Em Zeger e Liang (1986), as matrizes de correlação, $R_i(\alpha)$, são para cada UE, isto é, para cada y_i , de ordem $n_i \times n_i$. Isso significa, agora, que tanto a quantidade de observações (ao longo do tempo) quanto a própria matriz de correlação podem variar de UE para UE, permitindo, portanto, acomodar funções de *quase-verossimilhança* no nível da própria UE, para o processo de modelagem GEE. Consequentemente, a matriz de variância V_i será dada por

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} / \phi, \quad (2.47)$$

de modo que a matriz diagonal A_i , de ordem $n_i \times n_i$, engloba agora os elementos $g(\mu_{ij})$, função conhecida da média μ_i da i -ésima UE, posicionada na j -ésima posição.

As matrizes $R_i(\alpha)$ também são consideradas matrizes de correlação “working”, no sentido de que não necessariamente se espera que elas sejam especificadas corretamente. Desse modo, é possível obter estimativas consistentes, com variância assintoticamente consistente, ainda que $R_i(\alpha)$ esteja incorreta.

Como adaptação de S_k (estimador de quase-verossimilhança) para dados longitudinais, Zeger e Liang (1986) oferecem

$$\sum_{i=1}^K D_i^T V_i^{-1} S_i = 0, \quad (2.48)$$

versão parecida com aquela apresentada em Liang e Zeger (1986), com $S_i = y_i - \mu_i$, $\mu_i = (\mu_{i1}, \dots, \mu_{in_i})'$ e $D_i = \partial \mu_i / \partial \beta$. É possível notar, ainda, que, se $n_i = 1$, para todo i , o estimador de *quase-verossimilhança* da forma longitudinal (2.48) reduz-se ao estimador de *quase-verossimilhança* descrito em (2.46).

Pelo teorema 2 descrito em Liang e Zeger (1986), inclusive supondo as mesmas condições de regularidade, as expressões das matrizes de covariância (dos estimadores de *verossimilhança* e de *quase-verossimilhança*) exibem a mesma forma.

A solução dos estimadores GEE de *quase-verossimilhança* também pode ser obtida rigorosamente pelos mesmos procedimentos descritos em Liang e Zeger (1986) (método IRLS), quando do uso de funções de verossimilhança.

Além disso, sugere quatro possibilidades para R_i : i) a mais simples de todas, isto é, a matriz identidade I_{n_i} , indicando que as observações repetidas são não correlacionadas, ii) uma outra matriz $R_i(\alpha) = R(\alpha)$, em que se considera $n_i = n$, em situações específicas nas quais os tempos de observações são os mesmos para todos os indivíduos, iii) uma outra muito utilizada

em estudos longitudinais (principalmente em modelos de efeitos aleatórios), ao assumir que $[R_i]_{jk} = \alpha$, $j \neq k$ e, por último, iv) a matriz R_i pode assumir a forma

$$[R_i]_{jk} = \begin{cases} \alpha^{|t_{ij}-t_{ik}|}, & |t_{ij} - t_{ik}| \leq m \\ 0, & |t_{ij} - t_{ik}| > m \end{cases} \quad (2.49)$$

sendo t_{ij} e t_{ik} os tempos das observações j e k , respectivamente, de um determinado indivíduo i . Nesse caso, diz-se que a estrutura de correlação segue um processo estacionário.

Dois anos mais adiante, Zeger et al. (1988), em “Models for Longitudinal Data: A Generalized Estimating Equation Approach”, caracterizaram as abordagens de modelagem de modelos longitudinais em duas classes: modelos SS (*subject-specific*) e PA (*population-average*), tema a ser introduzido no Capítulo 6.

2.6 Apêndice

• Equação 2.8:

Com relação à equação (2.8), temos que a função log-verossimilhança é dada por

$$\ell_1(\beta, \rho) = \sum_{i=1}^K \left[y_{i1} Z_i' \beta - \log(1 + e^{Z_i \beta}) + \sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\} \right].$$

Por simplificação de exposição, dividiremos o somatório contido em $\ell_1(\beta, \rho)$ em duas partes:

- i) Parte A, isto é, $y_{i1} Z_i' \beta - \log(1 + e^{Z_i \beta})$;
- ii) Parte B, isto é, $\sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\}$.

Serão feitas derivadas parciais da parte A, $\frac{\partial \ell_A}{\partial \beta}$, e, em seguida, da parte B $\frac{\partial \ell_B}{\partial \beta}$. Ao final, essas derivadas parciais serão agregadas de forma que o resultado (2.8), $\frac{\partial \ell_1}{\partial \beta_u}$, seja alcançado.

Derivada parcial da Parte A:

$$\begin{aligned} \frac{\partial \ell_A}{\partial \beta_u} &= y_{i1} Z_{iu} - \frac{Z_{iu} e^{Z_i \beta}}{1 + e^{Z_i \beta}} \\ &= Z_{iu} \left(y_{i1} - \frac{e^{Z_i \beta}}{1 + e^{Z_i \beta}} \right) \\ &= Z_{iu} (y_{i1} - \pi_i). \end{aligned}$$

Derivada parcial da Parte B:

É importante lembrar que $\pi_{it} = \pi_i + \rho(Y_{i,t-1} - \pi_i)$, ou seja,

$$\pi_{it} = \pi_i + \rho Y_{i,t-1} - \rho \pi_i$$

$$\pi_{it} = \rho Y_{i,t-1} + (1 - \rho) \pi_i.$$

Uma vez que $\pi_i = \frac{e^{Z_i\beta}}{1+e^{Z_i\beta}}$, temos, então, que $1 - \pi_i = 1 - \frac{e^{Z_i\beta}}{1+e^{Z_i\beta}} = \frac{1+e^{Z_i\beta}-e^{Z_i\beta}}{1+e^{Z_i\beta}} = \frac{1}{1+e^{Z_i\beta}}$.

Logo, temos que π_{it} pode ser reescrito como

$$\pi_{it} = \rho Y_{i,t-1} + (1 - \rho) \frac{e^{Z_i\beta}}{1 + e^{Z_i\beta}},$$

de modo que é possível notar a relação de π_{it} e β .

Logo, a derivada parcial da Parte B será dada por

$$\begin{aligned} \frac{\partial \ell_B}{\partial \beta_u} &= \sum_{t=2}^{n_i} \left\{ y_{it} \frac{(\pi_{it})'}{\pi_{it}} + (1 - y_{it}) \frac{(-\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ y_{it} \frac{(\pi_{it})'}{\pi_{it}} + \frac{(-\pi_{it})'}{(1 - \pi_{it})} + y_{it} \frac{(\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ \frac{y_{it}(\pi_{it})'}{\pi_{it}} + \frac{(-\pi_{it})'}{(1 - \pi_{it})} + \frac{y_{it}(\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ y_{it}(\pi_{it})' \left(\frac{1}{\pi_{it}} + \frac{1}{1 - \pi_{it}} \right) + \frac{(-\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ y_{it}(\pi_{it})' \left(\frac{1 - \pi_{it} + \pi_{it}}{\pi_{it}(1 - \pi_{it})} \right) + \frac{(-\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ y_{it}(\pi_{it})' \left(\frac{1}{\pi_{it}(1 - \pi_{it})} \right) + \frac{(-\pi_{it})'}{(1 - \pi_{it})} \right\} \\ &= \sum_{t=2}^{n_i} \left\{ \frac{y_{it}(\pi_{it})'}{\pi_{it}(1 - \pi_{it})} + \frac{(-\pi_{it})'}{(1 - \pi_{it})} \right\}. \end{aligned}$$

Como relação, especificamente, à derivada $(\pi_{it})'$, temos que

$$\begin{aligned} (\pi_{it})' &= (1 - \rho) \frac{Z_{iu} e^{Z_i\beta} (1 + e^{Z_i\beta}) - e^{Z_i\beta} Z_{iu} e^{Z_i\beta}}{(1 + e^{Z_i\beta})^2} \\ &= (1 - \rho) \frac{Z_{iu} e^{Z_i\beta} [1 + e^{Z_i\beta} - e^{Z_i\beta}]}{(1 + e^{Z_i\beta})^2} \\ &= (1 - \rho) \frac{Z_{iu} e^{Z_i\beta}}{(1 + e^{Z_i\beta})^2} \\ &= Z_{iu} (1 - \rho) \frac{e^{Z_i\beta}}{(1 + e^{Z_i\beta})^2} \\ &= \frac{Z_{iu} (1 - \rho)}{(1 + e^{Z_i\beta})} \frac{e^{Z_i\beta}}{(1 + e^{Z_i\beta})} \end{aligned}$$

$$= \frac{Z_{iu}(1-\rho)}{(1+e^{Z_i\beta})} \pi_i.$$

Ao substituir a expressão de $(\pi_{it})' = \frac{Z_{iu}(1-\rho)}{(1+e^{Z_i\beta})} \pi_i$ em $\frac{\partial \ell_B}{\partial \beta}$, teremos

$$\begin{aligned} \frac{\partial \ell_B}{\partial \beta_u} &= \sum_{t=2}^{n_i} \left\{ \frac{y_{it} \frac{Z_{iu}(1-\rho)}{(1+e^{Z_i\beta})} \pi_i}{\pi_{it}(1-\pi_{it})} + \frac{-\frac{Z_{iu}(1-\rho)}{(1+e^{Z_i\beta})} \pi_i}{(1-\pi_{it})} \right\} \\ &= (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{y_{it} \frac{Z_{iu}}{(1+e^{Z_i\beta})} \pi_i}{\pi_{it}(1-\pi_{it})} + \frac{-\frac{Z_{iu}}{(1+e^{Z_i\beta})} \pi_i}{(1-\pi_{it})} \right\} \\ &= (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{\frac{y_{it} Z_{iu} \pi_i}{(1+e^{Z_i\beta})}}{\pi_{it}(1-\pi_{it})} + \frac{-\frac{Z_{iu} \pi_i}{(1+e^{Z_i\beta})}}{(1-\pi_{it})} \right\}. \end{aligned}$$

Uma vez que $1-\pi_i = \frac{1}{1+e^{Z_i\beta}}$, teremos, então, que

$$\frac{\partial \ell_B}{\partial \beta_u} = (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{y_{it} Z_{iu} \pi_i (1-\pi_i)}{\pi_{it}(1-\pi_{it})} - \frac{Z_{iu} \pi_i (1-\pi_i)}{(1-\pi_{it})} \right\}.$$

Ao considerarmos $\bar{\pi}_i = (1-\pi_i)$ e $\bar{\pi}_{it} = (1-\pi_{it})$, teremos

$$\begin{aligned} \frac{\partial \ell_B}{\partial \beta_u} &= (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{y_{it} Z_{iu} \pi_i \bar{\pi}_i}{\pi_{it} \bar{\pi}_{it}} - \frac{Z_{iu} \pi_i \bar{\pi}_i}{\bar{\pi}_{it}} \right\} \\ &= (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{y_{it} Z_{iu} \pi_i \bar{\pi}_i - \pi_{it} Z_{iu} \pi_i \bar{\pi}_i}{\pi_{it} \bar{\pi}_{it}} \right\} \\ &= (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{(Z_{iu} \pi_i \bar{\pi}_i)(y_{it} - \pi_{it})}{\pi_{it} \bar{\pi}_{it}} \right\}. \end{aligned}$$

Portanto, ao agregarmos as derivadas parciais das Partes A e B e aplicarmos o somatório para as K UEs, teremos

$$\frac{\partial \ell_1(\beta, \rho)}{\partial \beta_u} = \sum_{i=1}^K \left[Z_{iu}(y_{i1} - \pi_i) + (1-\rho) \sum_{t=2}^{n_i} \left\{ \frac{(Z_{iu} \pi_i \bar{\pi}_i)(y_{it} - \pi_{it})}{\pi_{it} \bar{\pi}_{it}} \right\} \right],$$

expressão do resultado (2.8).

• **Equação 2.9:**

A função log-verossimilhança é dada por

$$\ell_1(\beta, \rho) = \sum_{i=1}^K \left[y_{i1} Z_i' \beta - \log(1 + e^{Z_i \beta}) + \sum_{t=2}^{n_i} \{y_{it} \log \pi_{it} + (1 - y_{it}) \log(1 - \pi_{it})\} \right],$$

sendo que

$$\pi_i = \mathbb{P}(Y_{it} = 1 | Z_i)$$

e

$$\pi_{it} = \mathbb{E}(Y_{it}|Y_{i,t-1}, Z_i) = \pi_i + \rho(Y_{i,t-1} - \pi_i).$$

$$\pi_{it} = \mathbb{E}(Y_{it}|Y_{i,t-1}, Z_i) = \pi_i + \rho Y_{i,t-1} - \rho \pi_i$$

$$\pi_{it} = \mathbb{E}(Y_{it}|Y_{i,t-1}, Z_i) = \rho Y_{i,t-1} + (1 - \rho)\pi_i$$

É possível notar que, pelas premissas de Zeger et al. (1985), π_{it} é função de ρ . Dessa maneira, a derivada de π_{it} com relação a ρ é dada por $(\pi_{it})' = (Y_{i,t-1} - \pi_i)$.

É possível notar que, dentro dos colchetes, a derivada parcial de $y_{i1}Z_i'\beta - \log 1 + e^{Z_i\beta}$, com relação a ρ é igual a zero.

Logo, a derivada da função log-verossimilhança, com relação ao parâmetro ρ é dada por

$$\begin{aligned} \frac{\partial \ell_1(\beta, \rho)}{\partial \rho} &= \sum_{i=1}^K \sum_{t=2}^{n_i} [y_{it}(\log \pi_{it})' + (1 - y_{it})(\log(1 - \pi_{it}))'] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[y_{it} \frac{(\pi_{it})'}{\pi_{it}} + (1 - y_{it}) \frac{-(\pi_{it})'}{1 - \pi_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[y_{it} \frac{(\pi_{it})'}{\pi_{it}} + (1 - y_{it}) \frac{-(\pi_{it})'}{\bar{\pi}_{it}} \right]. \end{aligned}$$

Ao substituir $(\pi_{it})' = (Y_{i,t-1} - \pi_i)$, teremos,

$$\begin{aligned} \frac{\partial \ell_1(\beta, \rho)}{\partial \rho} &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[y_{it} \frac{(y_{i,t-1} - \pi_i)}{\pi_{it}} + (1 - y_{it}) \frac{-(y_{i,t-1} - \pi_i)}{\bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{\bar{\pi}_{it} y_{it} (y_{i,t-1} - \pi_i) - \pi_{it} (1 - y_{it}) (y_{i,t-1} - \pi_i)}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{\bar{\pi}_{it} y_{it} y_{i,t-1} - \bar{\pi}_{it} y_{it} \pi_i - (\pi_{it} - \pi_{it} y_{it}) (y_{i,t-1} - \pi_i)}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{\bar{\pi}_{it} y_{it} y_{i,t-1} - \bar{\pi}_{it} y_{it} \pi_i - [\pi_{it} y_{i,t-1} - \pi_{it} \pi_i - \pi_{it} y_{it} y_{i,t-1} + \pi_{it} y_{it} \pi_i]}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{\bar{\pi}_{it} y_{it} y_{i,t-1} - \bar{\pi}_{it} y_{it} \pi_i - \pi_{it} y_{i,t-1} + \pi_{it} \pi_i + \pi_{it} y_{it} y_{i,t-1} - \pi_{it} y_{it} \pi_i}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{(1 - \pi_{it}) y_{it} y_{i,t-1} - (1 - \pi_{it}) y_{it} \pi_i - \pi_{it} y_{i,t-1} + \pi_{it} \pi_i + \pi_{it} y_{it} y_{i,t-1} - \pi_{it} y_{it} \pi_i}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{y_{it} y_{i,t-1} - \pi_{it} y_{it} y_{i,t-1} - y_{it} \pi_i + \pi_{it} y_{it} \pi_i - \pi_{it} y_{i,t-1} + \pi_{it} \pi_i + \pi_{it} y_{it} y_{i,t-1} - \pi_{it} y_{it} \pi_i}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{y_{it} y_{i,t-1} - y_{it} \pi_i - \pi_{it} y_{i,t-1} + \pi_{it} \pi_i}{\pi_{it} \bar{\pi}_{it}} \right] \\ &= \sum_{i=1}^K \sum_{t=2}^{n_i} \left[\frac{(y_{it} - \pi_{it})(y_{i,t-1} - \pi_i)}{\pi_{it} \bar{\pi}_{it}} \right], \end{aligned}$$

em acordo com a expressão (2.9).

- **Equação 2.19:**

Em um contexto GLM, uma função densidade de probabilidade pode ser descrita, segundo Paula (2013, p. 4) diretamente sob a forma da família exponencial

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)],$$

em que $\mathbb{E}(Y_i) = \mu_i = b'(\theta_i)$, equivalente à expressão (2.17) e a $Var(Y_i) = \phi^{-1}V(\mu_i)$, semelhante à expressão (2.18) se considerarmos $V_i = V(\mu_i) = \frac{d\mu_i}{d\theta_i} = b''(\theta_i)$ e ϕ , um parâmetro de dispersão.

Uma outra característica dessa classe de modelos diz respeito à existência de uma função de ligação g , monótona e diferenciável, que conecta o valor esperado da resposta, μ_i , com o preditor linear, η_i , isto é,

$$g(\mu_i) = \eta_i,$$

com $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ representando um vetor de parâmetros desconhecidos e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ o vetor das variáveis explicativas.

A função log-verossimilhança de um GLM, com respostas independentes, é descrita sob a forma

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

No intuito de estimar $\boldsymbol{\beta}$, a função escore pode ser obtida por meio das sucessivas derivadas parciais, ou seja,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^n \phi \left\{ \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\ &= \sum_{i=1}^n \phi \left\{ y_i V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} - \mu_i V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\ &= \phi \sum_{i=1}^n (y_i - \mu_i) V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij}. \end{aligned}$$

É interessante notar que a conexão entre μ_i e η_i é feita via função de ligação, g , de acordo com a sua forma. Assim, μ_i é obtida pela inversa da função de ligação, isto é, $\mu_i = g^{-1}(\eta_i)$. Logo, a derivada $\frac{d\mu_i}{d\eta_i}$ não apresenta uma forma pré-determinada, uma vez que depende do tipo específico de função de ligação escolhida.

Com relação à forma matricial, a função escore assume a forma

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}).$$

Em Liang e Zeger (1986), o estimador $\hat{\beta}_I$ para β assume que as observações independentes de uma mesma UE são independentes. Assumindo, ainda, independência entre as K UEs, a solução para $\hat{\beta}_I$ consistiria em uma soma de K equações escore dada pela equação (2.19), isto é, dada por

$$U_I(\beta) = \sum_{i=1}^K X_i^T \Delta_i S_i = 0,$$

sendo $S_i = (Y_i - a'(\theta))$ e Δ_i uma matriz diagonal com elementos $\frac{d\theta_{it}}{d\eta_{it}}$.

A semelhança entre os métodos GLM e GEE é perceptível, explicitamente, entre os termos X_i^T e S_i , restando notar que Δ_i , matriz diagonal formada pelos elementos $\frac{d\theta_{it}}{d\eta_{it}}$ que, em um contexto GLM, seria dado pelas derivadas sucessivas $\frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i}$.

No caso da classe de modelos GLM, a função de variância, V_i , é escolhida logo quando determinada a função de verossimilhança. Por exemplo, se $Y_i \sim \text{Bernoulli}(\mu_i)$, então, $V_i = \mu_i(1 - \mu_i)$. Por outro lado, se $Y_i \sim \text{Poisson}(\lambda_i)$, então, $V_i = \lambda_i$ e assim por diante.

No caso da classe de modelos GEE, porém, especificamente em um contexto generalizado (sem a suposição de independência entre as observações de uma mesma UE), a função de variância, V_i , será estimada em conjunto com a matriz de correlação, $R(\alpha)$, isto é,

$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} / \phi,$$

sendo A_i uma matriz diagonal com elementos $a''(\theta_{it})$, que, pela expressão (2.18), equivale a uma das propriedades da família exponencial, a saber, que $\text{Var}(y_{it}) = a''(\theta_{it}) / \phi$.

- **Equação 2.22:**

Já em um contexto de GEE generalizado, isto é, com correlação entre medidas repetidas de uma mesma UE (e independência entre UEs), a solução para $\hat{\beta}_G$ é dada pela expressão (2.22), isto é, por

$$\sum_{i=1}^K D_i^T V_i^{-1} S_i = 0,$$

com $D_i = \frac{d\{a'_i(\theta)\}}{d\beta} = A_i \Delta_i X_i$. Comparativamente à classe de modelos GLM, o termo D_i encontraria relativa correspondência com as derivadas sucessivas $\frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j}$. Já o termo V_i^{-1} encontraria relativa correspondência com a derivada $\frac{d\theta_i}{d\mu_i}$ e, por fim, o termo S_i , com as derivadas $\frac{d\ell_i}{d\theta_i}$ e $\frac{\partial \eta_i}{\partial \beta_j}$.

3 O MODELO GLMM

3.1 Introdução

Neste capítulo será introduzida a classe de modelos *Generalized Linear Mixed Models* (GLMM), a qual pode ser entendida como uma extensão tanto da classe dos modelos LMM (*Linear Mixed Models*) quanto da classe de modelos GLM (*generalized linear models*).

Antes de entrarmos com detalhes nos modelos GLMM, algumas considerações devem ser feitas entre essas duas classes de modelos.

No que tange, inicialmente, à classe dos modelos LMM, essa consiste em uma classe de modelos que permite levar em consideração (e, portanto, modelar) efeitos aleatórios entre as UEs, assumindo distribuição normal para as variáveis respostas.

Os modelos LMM já permitiam análises com medidas repetidas, inclusive, sendo preferíveis aos modelos ANOVA (com medidas repetidas), uma vez que os modelos LMM permitem a acomodação tanto de dados balanceados quanto dados não balanceados.

Uma característica dos modelos LMM diz respeito ao fato de permitirem estimar efeitos aleatórios no nível das UEs. Fitzmaurice et al. (2011) reiteram, acerca da suposição implícita dos LMM, que

... some subset of the regression parameters vary randomly from one individual to another, thereby accounting for sources of natural heterogeneity in the population.” (p. 189)

Por conseguinte, a possibilidade de estimar efeitos aleatórios por parte dos modelos LMM também antecede a geração da classe de modelos GLMM.

Uma outra observação em relação aos modelos LMM é que a variável resposta admite apenas uma única distribuição, a saber, a distribuição normal. A título de ilustração, Jiang (2007), em seu prefácio, chama a atenção para o fato de os modelos LMM também serem conhecidos como *Gaussian (Linear) Mixed Models*.

O que, enfim, distingue a classe de modelos LMM da classe de modelos GLMM reside, particularmente, no fato de a variável resposta admitir, em um contexto GLMM, não apenas uma única distribuição (normal), mas todo um conjunto de distribuições de probabilidade, sejam discretas ou contínuas.

Contudo, também é possível afirmar que a classe de modelos GLMM também estende a classe de modelos GLM, a qual constitui uma família de modelos de regressão que permite

supor variáveis respostas não restritas à distribuição normal, ou seja, podem assumir a forma poisson, gama, exponencial etc. Brown e Prescott (2006) declaram que

A class of models known as generalized linear models (GLMs) is available for fitting fixed effects models to such non-normal data. These models can be further extend to fit mixed models and are then referred to as generalized linear mixed models (GLMMs). Random effects, random coefficients or covariance patterns can be included in a GLMM in much the same way as in normal mixed models, and again either balanced or unbalanced data can be analysed. (p. 107)

Nessa mesma linha de raciocínio, Breslow e Clayton (1993) alegam que:

The generalized linear model (GLM) (McCullagh and Nelder 1989) neatly synthesizes likelihood-based approaches to regression analysis for a variety of outcome measures. Several recent extensions of this useful theory involve models with random terms in the linear predictor. Such generalized linear mixed models (GLMMs) are useful for accommodating the overdispersion often observed among outcomes that nominally have binomial (Willians 1982) or Poisson (Breslow 1984) distributions; for modeling the dependence among outcomes variables inherent in longitudinal or repeated measures designs (Stiratelli, Laird, and Ware 1984; Zeger, Liang, and Albert 1988; and for producing shrinkage estimates in multiparameter problems, such as the construction of maps of small area disease rates (Clayton and Kaldor 1987; Manton et al. 1989). (p. 9)

Assim, de um modo geral, é possível concluir que a classe GLMM é extensão tanto da classe de modelos LMM, na medida em que acomoda outras distribuições (discretas ou contínuas) para a variável resposta, quanto da classe de modelos GLM, uma vez que permitem estimação de efeitos aleatórios específicos.

Em contrapartida, é possível afirmar que além da classe de modelos GLMM lidar com duas fontes de variabilidade (*entre-clusters* e *intra-cluster*), também oferece diferentes opções para a distribuição da variável resposta, características essas que os próprios modelos LMM não possuem.

Nas próximas seções serão abordadas, em detalhes, as características, propriedades e especificidades da classe de modelos GLMM (Seção 3.3) a partir do próprio artigo que a originou, a saber, o de Breslow e Clayton (1993), intitulado “Approximate Inference in Generalized Linear Mixed Models”. Mas, primeiramente, será introduzida a classe de modelos LMM (Seção 3.2), a qual antecede o surgimento da própria classe dos modelos GLMM, pelo artigo de Laird e Ware (1982), “Random-Effects Models for Longitudinal Data”, no intuito de colocar em perspectiva a evolução dessa classe de modelos.

3.2 O modelo LMM

O artigo de Laird e Ware (1982) subdivide-se em sete seções, mas somente as seguintes seções receberão uma maior atenção neste capítulo:

- modelos (Seção 2);
- estimação e inferência com dados de respostas medidas (Seção 3);
- o algoritmo EM e estimativas de máxima verossimilhança (Seção 4);
- estimação REML e cálculo (Seção 5);
- discussão (Seção 7).

É importante ressaltar que Laird e Ware (1982) desenvolveram a classe LMM de modelos por meio do procedimento em *dois estágios*.

No primeiro, a atenção estaria voltada para a apresentação da estrutura do modelo, de um modo geral, enquanto que, no segundo estágio, o foco estaria voltado para os parâmetros de efeitos aleatórios, os quais variam de UE para UE, segundo uma distribuição de probabilidades, com média e matriz de covariância.

De acordo com os autores, a estratégia de modelagem em *dois estágios*, em sua maioria, era comumente empregada, na literatura, em dois tipos de modelos:

- i) modelos de crescimento;
- ii) modelos de medidas repetidas.

O primeiro tipo de modelo realça as análises das variações no nível da UE a partir de uma perspectiva de desenvolvimento natural e ou de um processo de amadurecimento/envelhecimento característico da própria UE.

Já os modelos de medidas repetidas supõem que os efeitos no nível da UE não se alteram no decorrer do tempo de interesse. Nesse caso, seriam as *condições experimentais* que estariam sofrendo alterações ao longo do tempo (em razão de terem sido planejadas, ou não), de modo que os *efeitos experimentais* seriam responsáveis pelas variações no nível da UE.

Além disso, uma das alegações apresentadas por Laird e Ware (1982), acerca da necessidade dessa nova estratégia de modelagem, residiria no fato de que modelos multivariados com uma estrutura de covariância genérica encontravam barreiras para aplicações em um contexto de dados não balanceados. Como exemplo de dados não balanceados, os autores citam a incapacidade de se controlar completamente o âmbito no qual as mensurações são realizadas, no sentido de poder haver considerável variabilidade entre as UEs, seja com relação ao número de observações, seja com relação ao controle da periodicidade (*timing*) em que as observações eram realizadas.

Laird e Ware (1982) chamam a atenção, ainda, sobre os desenvolvimentos recentes de alguns métodos de estimação de modelos, os quais serão postos para discussão no próprio artigo, e que são centrais para a classe de modelos então propostos.

Dentre esses métodos, alguns descritos precisamente em Harville (1977) são diretamente responsáveis pelo desenvolvimento dos modelos GLMM.

Definem o vetor de respostas, $\mathbf{y}_i$, com distribuição normal multivariada com média $\boldsymbol{\mu}_i$ ($n_i \times 1$) e matriz de dispersão $\boldsymbol{\Sigma}$ de dimensão ($n_i \times n_i$), sendo n_i o número de observações da i -ésima UE, $i = 1, \dots, m$. Relembrem, ainda, que, em caso de dados estritamente balanceados, mas com dados faltantes, o método multivariado poderia, ainda, vir a ser aplicado.

Todavia, se medições individuais fossem realizadas em momentos arbitrários ou únicos ou se a dimensão da matriz de dispersão $\boldsymbol{\Sigma}$ fosse grande, a estratégia de utilizar o método multivariado tornar-se-ia pouco atraente, uma vez que poderia haver proliferação do número de parâmetros (de variância), além de não levar em consideração a possibilidade de estimar efeitos aleatórios.

Em seguida, os autores estabeleceram $\boldsymbol{\beta}$ como um vetor $p \times 1$ de parâmetros (populacionais) desconhecidos e $\mathbf{X}_i$ uma matriz de planejamento $n_i \times p$, $\mathbf{b}_i$ um vetor desconhecido de efeitos individuais e $\mathbf{Z}_i$ uma matriz de planejamento, $n_i \times k$.

1º Estágio:

Neste primeiro estágio, um modelo geral para a i -ésima UE, com $i = 1, \dots, m$, é descrito por meio da equação (3.1), levando em consideração as medidas repetidas ao longo de um determinado período de tempo.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i. \quad (3.1)$$

Os erros $\mathbf{e}_i$'s são independentes, distribuídos segundo a distribuição normal com média $\mathbf{0}$ e matriz de covariância $\mathbf{R}_i$ positiva definida, com dimensão $n_i \times n_i$, isto é, $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. Nesse estágio, tanto $\boldsymbol{\beta}$ quanto $\mathbf{b}_i$ são considerados como fixados.

2º Estágio:

Já no segundo estágio, os $\mathbf{b}_i$ são distribuídos (também) segundo a distribuição normal com média $\mathbf{0}$ e matriz de covariância positiva definida $\mathbf{D}$, de dimensão $k \times k$, isto é, $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$. Nesse estágio, somente $\boldsymbol{\beta}$ é tratado como sendo de efeitos fixos.

Do ponto de vista marginal, $\mathbf{y}_i$ são independentes e assumem distribuição normal com média $\mathbf{X}_i\boldsymbol{\beta}$ e matriz de variância-covariância $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \mathbf{R}_i$. Em outros termos, temos que

$$\mathbb{E}(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta} \quad (3.2)$$

e

$$\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \mathbf{R}_i. \quad (3.3)$$

As demonstrações de (3.2) e (3.3) encontram-se no apêndice deste capítulo (Seção 3.4).

Repare que, eventualmente, $\mathbf{R}_i = \sigma^2 \mathbf{I}$ (situação simplificadora) e que, nesse caso, $\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}$, refletindo o fato de que as respostas da i -ésima UE seriam independentes. É possível considerar, sem perda de generalidade, $\text{Var}(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{R}_i$.

3.2.1 Estimação e Inferência

Nesta seção, três situações serão analisadas:

- i) Quando os parâmetros de variância-covariância são conhecidos;
- ii) Quando os parâmetros de variância-covariância são desconhecidos;
- iii) Sobre os problemas envolvidos no processo de estimação.

i) Parâmetros de variância-covariância conhecidos

Em uma situação na qual todos os parâmetros de variância-covariância são conhecidos, e $\boldsymbol{\beta}$ é tratado como efeito fixo, as expressões para os efeitos populacionais e individuais assumem, respectivamente, a forma

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^m \mathbf{X}_i^T \mathbf{W}_i \mathbf{y}_i \quad (3.4)$$

e

$$\hat{\mathbf{b}}_i = \mathbf{D} \mathbf{Z}_i^T \mathbf{W}_i (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}), \quad (3.5)$$

com a suposição de existência da matriz inversa de $\mathbf{V}_i$, de modo que $\mathbf{W}_i = \mathbf{V}_i^{-1}$.

O estimador $\hat{\boldsymbol{\beta}}$, além de maximizar a função de verossimilhança da distribuição marginal dos dados, também é um estimador não viesado e de variância mínima.

O estimador $\hat{\mathbf{b}}_i$ não chega a ser um estimador de máxima verossimilhança, mas pode ser obtido a partir de uma extensão do teorema de Gauss-Markov descrito em Harville (1976), tendo sido aprimorado em Harville (1977). Além disso, a expressão de $\hat{\mathbf{b}}_i$ também pode ser obtida via procedimento bayesiano denominado *empirical bayes*, cuja finalidade consiste em encontrar uma distribuição *a priori* a partir dos dados (e não, conforme o procedimento bayesiano tradicional, pelo estabelecimento de uma distribuição *a priori* teórica e anterior à observação dos dados).

Uma vez que $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{b}}_i$ são, ambas, funções lineares de $\mathbf{y}$, suas expressões de erro padrão são descritas, respectivamente, por

$$\mathbb{V}ar(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^m \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \right)^{-1} \quad (3.6)$$

e

$$\mathbb{V}ar(\hat{\mathbf{b}}_i) = \mathbf{D} \mathbf{Z}_i^T \left\{ \mathbf{W}_i - \mathbf{W}_i \mathbf{X}_i \left(\sum_{i=1}^m \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \right)^{-1} \mathbf{X}_i^T \mathbf{W}_i \right\} \mathbf{Z}_i \mathbf{D}, \quad (3.7)$$

sendo que, se o objetivo consistir em avaliar o erro de estimação, $\hat{\mathbf{b}}_i - \mathbf{b}_i$, então, a expressão (3.7) deveria ser substituída por (3.7b), uma vez que (3.7) estaria subestimando a estimativa por não levar em consideração $\mathbb{V}ar(\mathbf{b}_i)$.

$$\mathbb{V}ar(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \mathbf{D} - \mathbf{D} \mathbf{Z}_i^T \mathbf{W}_i \mathbf{Z}_i \mathbf{D} + \mathbf{D} \mathbf{Z}_i^T \mathbf{W}_i \mathbf{X}_i \left(\sum_{i=1}^m \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i \right)^{-1} \mathbf{X}_i^T \mathbf{W}_i \mathbf{Z}_i \mathbf{D}. \quad (3.7b)$$

ii) Parâmetros de variância-covariância desconhecidos

Em uma situação na qual as matrizes de covariância são desconhecidas, mas uma estimativa de $\boldsymbol{\theta}$ (e, portanto, de $\mathbf{R}_i$ e $\mathbf{D}$) é possível, então, segundo Laird e Ware (1982), é razoável conceber

$$\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\mathbf{D}} \mathbf{Z}_i^T + \hat{\mathbf{R}}_i = \hat{\mathbf{W}}_i^{-1}, \quad (3.8)$$

de modo que estimativas para $\boldsymbol{\beta}$ e $\mathbf{b}_i$ possam, ainda, ser obtidas pelas equações (3.4) e (3.5), substituindo cada $\mathbf{W}_i$ por $\hat{\mathbf{W}}_i$, as quais passariam a ser descritas por $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$.

Esse tipo de problema, a saber, de estimação conjunta de parâmetros (no caso, de vetores paramétricos $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$) pela maximização de verossimilhança conjunta é bem comum na área da Estatística, no sentido de que, a partir da obtenção de uma das estimativas, é possível obter uma outra. Em outras palavras, é possível afirmar que, em situações nas quais se estivesse em busca de estimativas de máxima verossimilhança para $(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M)$, essas também seriam passíveis de serem obtidas via $\hat{\boldsymbol{\beta}}_M = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_M)$, ou seja, a partir de $\hat{\boldsymbol{\theta}}_M$ poder-se-ia obter uma estimativa para $\hat{\boldsymbol{\beta}}_M$.

Na mesma linha de raciocínio, se se estivesse em busca de uma estimativa de máxima verossimilhança para $\mathbf{b}$, $\mathbf{b}^T = (\mathbf{b}_1^T, \dots, \mathbf{b}_m^T)$, isto é, $\hat{\mathbf{b}}_M$, então, essa estimativa poderia ser obtida a partir de estimativas de máxima verossimilhança de outras quantidades ou, em outros termos,

$$\begin{aligned} \hat{\mathbf{b}}_M &= \mathbb{E}(\mathbf{b} \mid \mathbf{y}, \hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\theta}}_M) \\ &= \hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M), \end{aligned} \quad (3.9)$$

o que, conforme mencionado anteriormente, configura uma estimativa *empirical bayes* para $\mathbf{b}$, com $\boldsymbol{\theta}$ estimado via máxima verossimilhança.

Já com relação às estimativas de erro padrão de $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$, Laird e Ware (1982) sugerem substituir $\hat{\boldsymbol{\theta}}$ nas expressões de $\text{Var}(\boldsymbol{\beta})$ e $\text{Var}(\mathbf{b}_i)$, isto é, nas expressões (3.6) e (3.7), lembrando que o vetor $\boldsymbol{\theta}$ (constituído basicamente por parâmetros de variância e covariância) abrange componentes pertencentes a $\mathbf{R}_i$ e $\mathbf{D}$.

Dessa maneira, uma vez obtida uma estimativa de $\hat{\boldsymbol{\theta}}$, a obtenção tanto de $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ quanto de $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$ (bem como as estimativas de seus respectivos erros padrão) poderia ser alcançada, aparentemente, sem grandes problemas. Presumivelmente, o grande desafio pareceria incidir sobre os cálculos de estimação de $\boldsymbol{\theta}$, assunto da próxima subseção.

iii) Estimando a matriz de covariância

É importante destacar que os problemas de estimação dessa nova classe de modelos (LMM) envolvem questões metodológicas, práticas e técnicas de algoritmos, tais como cálculos, velocidade de convergência, vantagens e desvantagens relativas de escolha de algoritmos etc.

Apesar dos próprios autores descreverem o texto de Harville (1977) como uma revisão do “estado da arte” acerca de vários tipos de estimadores, bem como seus aspectos computacionais, decidem concentrar esforços sobre dois estimadores concorrentes, a saber,

- a) máxima verossimilhança (ML), isto é, $\hat{\boldsymbol{\theta}}_M$;
- b) máxima verossimilhança restrita (REML), $\hat{\boldsymbol{\theta}}_R$,

sendo que ambos serão estimados via algoritmo E-M (*Expectation-Maximization*).

Esse algoritmo, de acordo com Hogg, McKean e Craig (2005, p. 359), tem por finalidade encontrar estimativas de máxima verossimilhança ao implementar iterativamente dois passos, sendo que, no primeiro deles, calcula-se a esperança do log-verossimilhança e, no segundo passo, determina-se o valor do parâmetro que maximiza a esperança do log-verossimilhança determinada no primeiro passo. De modo mais ilustrativo, temos que

- 1º passo (**E-step**): Calcule $\mathbb{E}_{f(z|x, \theta^{(t)})}[\log f(x, z|\theta)]$;
- 2º passo (**M-step**): $\theta^{(t+1)} = \arg \max_{\theta} \mathbb{E}_{f(z|x, \theta^{(t)})}[\log f(x, z|\theta)]$.

Hogg et al. (2005, p. 360) asseguram ainda que, sob fortes suposições, a estimativa do parâmetro converge, em probabilidade, para uma estimativa de máxima verossimilhança na medida em que o número de iterações tende ao infinito.

No caso, o algoritmo E-M permitiria estimar quantidades tais como $\boldsymbol{\beta}(\hat{\boldsymbol{\theta}})$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$ e $\hat{\boldsymbol{\theta}}$, seja via $\hat{\boldsymbol{\theta}}_M$ ou via $\hat{\boldsymbol{\theta}}_R$.

A Seção 4 destina-se, principalmente, a abordar o cálculo de $\boldsymbol{\beta}(\hat{\boldsymbol{\theta}}_M)$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$ e $\hat{\boldsymbol{\theta}}_M$ via algoritmo E-M. Já na Seção 5, os autores abordaram, de um ponto de vista bayesiano, o método da máxima verossimilhança restrita, REML, e, conseqüentemente, a problemática da estimação de $\boldsymbol{\beta}(\hat{\boldsymbol{\theta}}_R)$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$ e $\hat{\boldsymbol{\theta}}_R$.

3.2.2 Usando o Algoritmo E-M para Estimativas de Máxima Verossimilhança

Laird e Ware (1982) relembam que Dempster et al. (1977) já tinham constatado que vários dos algoritmos iterativos que se destinavam a avaliar estimativas de verossimilhança constituíam, na realidade, casos particulares de um algoritmo bem mais geral, denominado algoritmo E-M, amplamente empregado em problemas de bases com dados faltantes.

Além disso, Laird e Ware (1982) observaram que Laird (1982) havia mostrado que, em um contexto no qual um vetor $\boldsymbol{\theta}$ de componentes de variância ($\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i \times n_i}$ e $\mathbf{D}$ é matriz diagonal), o algoritmo E-M seria equivalente ao algoritmo capaz de obter estimativas de máxima verossimilhança de Henderson, conforme descrito em Harville (1977).

Assim, ao colocar o problema de dados faltantes em um contexto de dados longitudinais, estimativas de máxima verossimilhança para $\mathbf{b}_i$ e $\mathbf{e}_i$, com formas fechadas, também seriam passíveis de serem obtidas por meio de suas formas quadráticas, isto é, por meio de $\mathbf{b}_i \mathbf{b}_i^T$ e $\mathbf{e}_i^T \mathbf{e}_i$, respectivamente. Por exemplo, se a matriz $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i \times n_i}$, com suposição de independência entre os $\mathbf{e}_i$, e $\mathbf{D}$ for uma matriz positiva semidefinida $k \times k$, as estatísticas suficientes

$$\hat{\sigma}^2 = \sum_{i=1}^m \mathbf{e}_i^T \mathbf{e}_i / \sum_{i=1}^m n_i = t_1 / \sum_{i=1}^m n_i \quad (3.10)$$

e

$$\hat{\mathbf{D}} = \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T / m = \mathbf{t}_2 / m, \quad (3.11)$$

poderiam ser utilizadas.

A existência de uma estimativa (inicial) $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ tornaria viável a obtenção das estimativas $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$ de $\boldsymbol{\beta}$ e $\mathbf{b}_i$, respectivamente, de forma a permitir que as “estatísticas suficientes estimadas”

$$\begin{aligned}\hat{t}_1 &= \mathbb{E} \left(\sum_{i=1}^m \mathbf{e}_i^T \mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}} \right) \\ &= \sum_{i=1}^m \left[\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}})^T \hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}}) + \text{tr} \mathbb{V}\text{ar}\{\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\} \right] \quad (3.12)\end{aligned}$$

e

$$\begin{aligned}\hat{t}_2 &= \mathbb{E} \left(\sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T \mid \mathbf{y}_i, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}} \right) \\ &= \sum_{i=1}^m \left[\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}) \hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})^T + \mathbb{V}\text{ar}\{\mathbf{b}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\} \right], \quad (3.13)\end{aligned}$$

fossem atualizadas.

Assim, Laird e Ware (1982) salientam que ao atribuir um valor inicial para $\hat{\boldsymbol{\theta}}$, obter-se-ia também uma estimativa $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$, que, se incorporada iterativamente em (3.12) e (3.13) (passo *E-step*) e, por conseguinte, também em (3.10) e (3.11) (passo *M-step*), uma estimativa de máxima verossimilhança, $\hat{\boldsymbol{\theta}}_M$, seria obtida de forma que, quando alcançada a convergência, as estimativas $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_M)$ e $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$ também seriam resultantes desse processo.

Com relação a (3.12), Laird e Ware (1982) observam que $\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}}) = \mathbb{E}(\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}) = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) - \mathbf{Z}_i \hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$, além de chamarem a atenção, também, para o fato de que o problema de convergência lenta pode ser grave, no contexto dessa estratégia de modelagem, se a estimativa de máxima verossimilhança estiver em uma região próxima dos limites do espaço paramétrico.

Por fim, os autores finalizam essa seção ressaltando que utilizam o algoritmo E-M para finalidades distintas, isto é, não para lidar com dados faltantes, mas, de outro modo, para estimar parâmetros não observáveis (de efeitos aleatórios).

3.2.3 Estimação REML e Cálculo

Conforme Fitzmaurice et al. (2011, p. 179) descrevem em detalhes, o estimador de máxima verossimilhança, $\hat{\sigma}_M^2$, para pequenas amostras é viesado, mas, por outro lado, esse viés é removido ao realizar um “ajuste” em seu denominador (dividir pelos graus de liberdade dos resíduos). Esse estimador diferenciado (não viesado) é conhecido como estimador de máxima verossimilhança restrita (REML). Nesse sentido, estimativas de componentes de variância pelo

método da máxima verossimilhança restrita, $\hat{\boldsymbol{\theta}}_R$, são, geralmente, preferíveis quando obtidos pelo método da máxima verossimilhança, $\hat{\boldsymbol{\theta}}_M$.

Nesta seção, uma análise das estimativas bayesianas via REML, bem como um método unificado de estimação de $\boldsymbol{\beta}$, $\mathbf{b}$ e $\boldsymbol{\theta}$, será abordada.

O modelo em *dois estágios*, conforme sugerido pelos autores, ainda permanece caracterizado como tal, porém, em uma perspectiva bayesiana. As expressões descritas a seguir em (3.14) e (3.15) podem ser conferidas no apêndice do Capítulo 3 (Seção 3.4).

O primeiro estágio permanece praticamente inalterado, com a observação de que, condicionado em $\boldsymbol{\beta}$ e $\mathbf{b}_i$, a variável resposta $\mathbf{y}_i$ possui distribuição normal com média $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ e matriz de variância $\mathbf{R}_i$ ou seja,

$$\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \mathbf{R}_i). \quad (3.14)$$

Já no segundo estágio, $\boldsymbol{\beta}$ e $\mathbf{b}_i$ são variáveis aleatórias independentes, normalmente distribuídas, ambas com média $\mathbf{0}$ e matrizes de variância $\text{Var}(\boldsymbol{\beta}) = \boldsymbol{\Gamma}$ e $\text{Var}(\mathbf{b}_i) = \mathbf{D}$, sendo que a independência entre $\boldsymbol{\beta}$ e $\mathbf{b}_i$ implica $\text{Cov}(\boldsymbol{\beta}, \mathbf{b}_i) = \mathbf{0}$. Marginalmente, temos que

$$\mathbf{y}_i \sim N(\mathbf{0}, \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i^T + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \mathbf{R}_i), \quad (3.15)$$

com a mesma lógica de vetor de parâmetros desconhecidos $\boldsymbol{\theta}$ em $\mathbf{D}$ e $\mathbf{R}_i$.

Se $\boldsymbol{\theta}$ e $\boldsymbol{\Gamma}$ fossem conhecidos, estimativas bayesianas para $\boldsymbol{\beta}$ e $\mathbf{b}$ poderiam ser obtidas por meio de suas distribuições *a posteriori*, precisamente por meio do valor esperado (ou da média) da distribuição *a posteriori*, dados $\mathbf{y}$, $\boldsymbol{\theta}$ e $\boldsymbol{\Gamma}$. Mas, no caso de $\boldsymbol{\theta}$ e $\boldsymbol{\Gamma}$ desconhecidos, uma estratégia possível para $\boldsymbol{\beta}$ e $\mathbf{b}$ consistiria em fazer uso do procedimento *empirical bayes*, substituindo $\boldsymbol{\theta}$ e $\boldsymbol{\Gamma}$ pelas suas respectivas estimativas, a partir da maximização da verossimilhança marginal (normal) de $\mathbf{y}_i$, integrando sobre $\boldsymbol{\beta}$ e $\mathbf{b}$.

Entretanto, frequentemente, não se dispõe das informações sobre $\boldsymbol{\Gamma}$, isto é, sobre as informações que permitiriam modelar a variabilidade de certas características populacionais $\boldsymbol{\beta}$.

Apesar disso, uma estratégia capaz de lidar com essa limitação consiste em estabelecer $\boldsymbol{\Gamma}^{-1} = \mathbf{0}$, isto é, em supor uma distribuição *a priori* pouco informativa acerca de $\boldsymbol{\beta}$, o que, em outras palavras, expressa, em termos matemáticos, o baixo nível de informação sobre tal vetor paramétrico.

Com base nessa proposta, seria possível obter uma estimativa $\hat{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$, ao maximizar a verossimilhança marginal limite ($\boldsymbol{\Gamma}^{-1} \rightarrow \mathbf{0}$) de $\boldsymbol{\theta}$, dado $\mathbf{y}$, expediente de grande valia, uma vez que, segundo os autores, Harville (1976) provou que essa função de verossimilhança limite equivale à expressão de verossimilhança dos estimadores REML, fato que permite colocar a

questão em outros termos, a saber, que as estimativas de $\boldsymbol{\theta}$, pelos procedimentos bayesianos ora descritos, também seriam estimativas REML.

Dessa maneira, as expressões para as médias das distribuições *a posteriori* seriam descritas por

$$\mathbb{E}(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\Gamma}^{-1} = \mathbf{0}, \hat{\boldsymbol{\theta}}_R) \quad (3.16)$$

e

$$\mathbb{E}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\Gamma}^{-1} = \mathbf{0}, \hat{\boldsymbol{\theta}}_R) = \mathbb{E}(\mathbf{b}_i \mid \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R), \hat{\boldsymbol{\theta}}_R), \quad (3.17)$$

que, nesse caso, seriam equivalentes a $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R)$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}_R)$, respectivamente.

Desse modo, ao fazer uso do procedimento *empirical bayes* com a finalidade de se obter estimativas REML, tornaria, então, possível obter, $\hat{\boldsymbol{\theta}}_R$, $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R)$ e $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$, via algoritmo E-M.

Dessa forma, uma combinação dos métodos de máxima verossimilhança com procedimentos bayesianos e algoritmo E-M configura os processos de estimação dessa nova classe de modelos.

A seguir, será analisada uma nova classe de modelos, a saber, uma classe de modelos que permite estimar efeitos aleatórios, não restrita única e exclusivamente à distribuição normal para a variável resposta.

3.3 O modelo GLMM

Nesta seção, será introduzida a classe de modelos GLMM, propiciada pelo artigo de Breslow e Clayton (1993). Essa classe de modelos expande a classe de modelos da seção anterior (LMM), de modo que a variável resposta possa admitir outras distribuições de probabilidade (que não a normal), podendo, além disso, admitir, também, o caso discreto.

Na medida em que passam a ser admitidas mais possibilidades para a variável resposta (distribuições discretas e contínuas), tanto a função de verossimilhança quanto a função de quase-verossimilhança voltam a assumir um papel de destaque nessa nova classe de modelos.

Isso porque, apesar dos modelos GLMM poderem ser estimados com funções de verossimilhança, os métodos de estimação dessa classe de modelos, em sua origem, surgiram para estimar modelos baseados, inicialmente, em funções de quase-verossimilhança.

Mais especificamente, Breslow e Clayton (1993) desenvolveram uma metodologia (PQL – *Penalized Quasilikelihood*) por meio da qual a aproximação da função de quase-verossimilhança via método de Laplace tornaria possível a estimação dessa classe de modelos.

Ao contrário do que se possa imaginar, o problema de estimação dessa classe de modelos (GEE) via função de verossimilhança (ou log-verossimilhança) não é trivial. De um ponto de vista puramente técnico, o processo de maximização da função de verossimilhança, em um contexto de GLMM, envolve, em geral, um número considerável de integrais, de modo que a própria obtenção de uma forma fechada para sua expressão se torna praticamente inviável. Em termos operacionais, o interesse reside em tornar possível o cálculo da integral (função de verossimilhança)

$$L(\boldsymbol{\beta}, \mathbf{D}, \phi) = \prod_{i=1}^N \int_{\mathbb{R}^k} \prod_{j=1}^{n_i} f_i(y_{ij} | \mathbf{b}_i; \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | \mathbf{D}) d\mathbf{b}_i. \quad (3.18)$$

De um ponto de vista intuitivo, a expressão (3.18) envolve uma função densidade de probabilidade $f_i(y_{ij} | \mathbf{b}_i; \boldsymbol{\beta}, \phi)$ para a variável resposta de uma determinada (*i*-ésima) UE, condicionada em seus efeitos aleatórios, para dados parâmetros de efeitos fixos $\boldsymbol{\beta}$ e de dispersão, ϕ , sendo $\mathbf{b}$ um vetor *k-dimensional*.

McCulloch e Searle (2001, p. 225), a título de ilustração, apresentam um exemplo originalmente descrito em Diggle et al. (1994) sobre um modelo longitudinal para o número de ataques epiléticos de pacientes.

Nesse caso, a variável resposta do modelo, y_{ij} , é o próprio número de ataques epiléticos do *i*-ésimo paciente ($i = 1, \dots, m$), no *j*-ésimo período ($j = 1, \dots, n_i$). Uma vez que a variável resposta é de contagem, a distribuição de *Poisson*(μ_{ij}) pode ser utilizada com função de ligação “log” para conectar a média da variável resposta, μ_{ij} , com o preditor linear, $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i$, de modo que $\log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i$. Poder-se-ia, ainda, supor distribuição normal para os efeitos aleatórios u_i 's dos pacientes, com média zero e variância σ_u^2 , ou seja, $u_i \sim N(0, \sigma_u^2)$. Essas informações caracterizam um modelo GLMM com resposta de Poisson cuja função log-verossimilhança assume a forma

$$\ell(\boldsymbol{\beta}, \mathbf{D}, \phi) = \log \left(\prod_{i=1}^m \int_{-\infty}^{+\infty} \prod_{j=1}^{n_i} \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}u_i^2} du_i \right), \quad (3.19)$$

tornando, assim, mais evidente o problema simplificação dessa expressão no intuito de obter estimativas que maximizam $\ell(\boldsymbol{\beta}, \mathbf{D}, \phi)$ por meio de uma forma fechada.

Apesar do problema de obtenção de uma forma fechada, vários métodos numericamente intensivos para a aproximação de (3.18), isto é, da função de verossimilhança, já foram desenvolvidos para a estimação de modelos GLMM. Dentre eles, Agresti (2015, p. 313)

destaca: i) Quadratura de Gauss-Hermite, ii) Aproximações de Laplace e iii) Métodos de Monte Carlo.

Somente com relação aos métodos de Monte Carlo, Agresti (2015, p. 313) ressalta que são várias as abordagens: i) com cadeias de Markov (MCMC), ii) Monte Carlo com algoritmo de Newton-Raphson e iii) Monte Carlo com algoritmo E-M.

Além disso, Wolfinger e O'connell (1993) também haviam criado (publicado em setembro de 1993, enquanto que a publicação de Breslow e Clayton data de março do mesmo ano) uma metodologia alternativa para estimar modelos GLMM com base na função de pseudoverossimilhança.

Capanu, Gönen e Begg (2013) puderam aperfeiçoar uma metodologia híbrida desenvolvida anteriormente, em Capanu e Begg (2011) e também baseada na função pseudoverossimilhança, com a substituição, no 3º passo (do algoritmo), das estimativas de pseudoverossimilhança por estimativas oriundas da aproximação de Laplace.

De um modo geral, existe já um certo número de métodos de estimação da classe de modelos GLMM, sendo, inclusive, essa, uma área em evolução.

É importante ressaltar ainda que o fato do método de estimação em Breslow e Clayton (1993) estar baseado na aproximação da função de quase-verossimilhança não constitui, de forma alguma, uma espécie de desvantagem. Essa função (quase-verossimilhança) é amplamente utilizada (inclusive em outras classes de modelos, tais como GEE e GLM), com McCulloch e Searle (2001, p. 232) destacando como atratividade sua capacidade de fornecer estimadores bastante eficientes sem a necessidade de conhecer e/ou supor uma distribuição para a variável resposta.

Quando Breslow e Clayton (1993) abordaram o método PQL de estimação, no contexto de GLMM, os autores acabaram por reconhecer que não se tratava de uma novidade em si, mas, que, de certa forma, esse método não havia, ainda, sido considerado em um nível de generalidade então proposto pelos autores.

A seguir, serão elaboradas e analisadas as condições e situações específicas para a utilização da função de quase-verossimilhança no que diz respeito ao método de estimação PQL para a classe de modelos GLMM.

3.3.1 A Função de Quase-verossimilhança

Inicialmente, será feita uma breve introdução ao método de aproximação de Laplace, também abordada em Breslow e Clayton (1993).

Quando, por exemplo, uma determinada integral (p. ex., função de verossimilhança) torna-se de difícil obtenção, o procedimento de aproximação tende a ser uma alternativa.

Suponha que se deseja aproximar uma integral sob a forma

$$\int e^{-q(\mathbf{b})} d\mathbf{b}, \quad (3.20)$$

sendo $q(\cdot)$ uma função tal que, quando avaliada em $\mathbf{b} = \hat{\mathbf{b}}$, atinge seu ponto de mínimo, com $q'(\hat{\mathbf{b}}) = 0$ e $q''(\hat{\mathbf{b}}) > 0$. Se for realizada uma expansão de Taylor de segunda ordem de $q(\mathbf{b})$, teremos

$$\begin{aligned} q(\mathbf{b}) &\approx q(\hat{\mathbf{b}}) + q'(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})q''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}) \\ &\approx q(\hat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})q''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}), \end{aligned} \quad (3.21)$$

uma vez que $q'(\hat{\mathbf{b}}) = 0$, tendo $q''(\hat{\mathbf{b}})$ como a matriz das derivadas de segunda ordem (matriz Hessiana), avaliada em $\hat{\mathbf{b}}$.

Ao combinar (3.20) e (3.21), obtém-se o resultado aproximado

$$\int e^{-q(\mathbf{b})} d\mathbf{b} \approx (2\pi)^{k/2} | -q''(\hat{\mathbf{b}}) |^{-1/2} e^{-q(\hat{\mathbf{b}})}, \quad (3.22)$$

conhecido como aproximação de Laplace para o caso *k-dimensional*. Há autores (Jiang (2007, p. 127)) que descrevem a aproximação de Laplace com a substituição de $(2\pi)^{k/2}$ por uma constante, c , a qual dependeria apenas da dimensão da integral (no caso, k).

Esse método de aproximação é bastante utilizado na Estatística Bayesiana e pode ser encontrado em vários textos técnicos. Em Ghosh, Delampady e Samanta (2006, p. 113), um resultado mais geral desse método (com um integrando mais complexo) pode ser encontrado com aplicação, inclusive, em uma situação semelhante (também de Estatística Bayesiana). Molenberghs e Verbeke (2005, p. 269) relatam, ainda, que a aproximação de Laplace será exata quando $q(\mathbf{b})$ for uma função quadrática de $\mathbf{b}$.

Contudo, é importante destacar que cada uma das integrais em (3.19) é proporcional a uma integral descrita sob a forma (3.20), uma vez que, nessa classe de modelos GLMM, a função densidade de probabilidade (da variável resposta) pertence à família exponencial. A título de ilustração, e supondo distribuição normal para os efeitos aleatórios, Sun (2011, p. 9) descreve a expressão (3.18) também sob a forma

$$L(\boldsymbol{\beta}, \mathbf{D}, \phi) = \prod_i \int_{-\infty}^{+\infty} e^{\Sigma_j [(y_i y_{ij} - b(\gamma_{ij}))/a(\phi) - c(y_i, \phi)]} \frac{e^{-\frac{u_i^2}{2\sigma_u^2}}}{\sqrt{2\pi\sigma_u^2}} du_i. \quad (3.23)$$

Logo, é possível notar que $L(\boldsymbol{\beta}, \mathbf{D}, \phi)$ pode, na realidade, ser reescrita como uma integral do tipo $\int e^{-q(\mathbf{b})} d\mathbf{b}$, tornando possível utilizar a aproximação de Laplace mais adiante.

Conforme descrito na Seção 2.5, é bastante comum, em modelos baseados na função de quase-verossimilhança, a utilização da função quase-desvio, sendo essa uma simples função (ou transformação) daquela. Isso, porque, no processo de estimação de $\hat{\boldsymbol{\beta}}$ (em um contexto GLM e, portanto, não para dados longitudinais), a utilização da função quase-desvio, $d(y; \mu)$, na função escore, $\mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ produz uma metodologia de estimação denominada função quase-escore, versão ligeiramente diferenciada da função escore (McCullagh & Nelder, 1989, p. 327).

A função de quase-verossimilhança, $Q(\mu, y)$, bem como a função quase-desvio, $d(y; \mu)$, encontram-se definidas, respectivamente, nas expressões (2.32) e (2.34), as quais são dadas, respectivamente, pelas expressões $Q(\mu, y) = \int_y^\mu \frac{y-t}{\sigma^2 V(t)} dt$ e $d(y; \mu) = -2\sigma^2 Q(\mu, y)$.

Em Breslow e Clayton (1993), a função de quase-verossimilhança a ser utilizada para estimar $(\boldsymbol{\beta}, \boldsymbol{\theta})$ é dada pela relação de proporcionalidade

$$e^{ql(\boldsymbol{\beta}, \boldsymbol{\theta})} \propto |\mathbf{D}|^{-1/2} \int \exp \left[-\frac{1}{2\phi} \sum_{i=1}^n d_i(y_i; \mu_i) - \frac{1}{2} \mathbf{b}' \mathbf{D}^{-1} \mathbf{b} \right] d\mathbf{b}, \quad (3.24)$$

sendo $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$ o termo para expressar a função de quase-verossimilhança e a função quase-desvio dada por $d_i(y; \mu) = -2 \int_y^\mu \frac{y-u}{a_i v(u)} du$.

O termo a_i no denominador de $d_i(y; \mu)$ refere-se, segundo os autores, a uma constante conhecida (podendo ser utilizada como recíproca de um denominador binomial). O vetor $\mathbf{b}$ admite a forma de uma função quadrática de $\boldsymbol{\beta}$, fato que implica, conforme mencionado anteriormente, em uma aproximação exata de Laplace e $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$ representa a matriz de covariância de $\mathbf{b}$, a qual possui distribuição normal com média zero.

Breslow (2004, p. 4) chama, ainda, a atenção para o fato de que se, em (3.24), y_i possui distribuição normal e se a função de ligação identidade, $g(\cdot)$, é utilizada, então, a integral em (3.24) poderia ser avaliada em uma forma fechada. Mas, em condições diferentes, a maximização da expressão já passa a exigir métodos numéricos.

Com relação à função quase-desvio, d_i , uma análise mais pormenorizada sobre algumas de suas características e propriedades encontra-se no Apêndice deste capítulo (Seção 3.4).

Contudo, em termos comparativos, os desafios enfrentados por Breslow e Clayton (1993) para o desenvolvimento de um método de estimação de parâmetros para a classe de modelos GLMM se diferenciam daqueles abordados no início da Seção 3.3 acerca da expressão $L(\boldsymbol{\beta}, \mathbf{D}, \phi)$ em (3.18), fundamentalmente por uma razão: a escolha foi pela função de quase-verossimilhança.

3.3.2 O método PQL

Breslow e Clayton (1993), na prática, assumem um modelo hierárquico para uma variável resposta da i -ésima UE, isto é, y_i , por meio dos vetores $\mathbf{x}_i$ e $\mathbf{y}_i$ de variáveis explanatórias associadas com efeitos fixos e aleatórios.

Ao considerar a expressão (3.24) sob a forma $|\mathbf{D}|^{-1/2} \int e^{-q(\mathbf{b})} d\mathbf{b}$, torna-se possível aplicar o método de Laplace no intuito de aproximar a integral. Assim, temos que

$$|\mathbf{D}|^{-1/2} \int e^{-q(\mathbf{b})} d\mathbf{b} \approx |\mathbf{D}|^{-1/2} c |q''(\tilde{\mathbf{b}})|^{-1/2} e^{-q(\tilde{\mathbf{b}})}, \quad (3.25)$$

tal que $\tilde{\mathbf{b}}$ minimiza $q(\mathbf{b})$. O logaritmo dessa expressão, $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$, assume a forma aproximada (ignorando a constante c)

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{D}| - \frac{1}{2} \log |q''(\tilde{\mathbf{b}})| - q(\tilde{\mathbf{b}}). \quad (3.26)$$

Apenas para fins de simplificação, consideraremos, a partir de (3.24), que

$$q(\mathbf{b}) = \frac{1}{2} \left(\sum_{i=1}^n d_i + \mathbf{b}' \mathbf{D}^{-1} \mathbf{b} \right), \quad (3.27)$$

com a observação de que o parâmetro de dispersão ϕ passou a ser incorporado por d_i , precisamente em seu denominador, isto é,

$$d_i = -2 \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi a_i v(t)} dt. \quad (3.28)$$

Para utilizar o método de Laplace, é, conforme a própria expressão de $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$, imprescindível o conhecimento de $q''(\tilde{\mathbf{b}})$. Isso envolve necessariamente a obtenção da derivada primeira $q'(\mathbf{b})$, procedimento capaz de fornecer uma possível solução para $\tilde{\mathbf{b}}$, uma vez que essa solução é passível de ser alcançada a partir de $q'(\mathbf{b}) = 0$.

Nessa linha de raciocínio, $\tilde{\mathbf{b}}$, obtido por meio de $q'(\mathbf{b}) = 0$ é dado por

$$\mathbf{D}^{-1} \mathbf{b} - \sum_{i=1}^n \frac{y_i - \mu_i}{\phi a_i v(\mu_i)} \frac{z_i}{g'(\mu_i)} = 0, \quad (3.29)$$

enquanto que, com relação à derivada segunda,

$$q''(\mathbf{b}) = \sum_{i=1}^n \frac{z_i z_i'}{\phi a_i v(\mu_i) \{g'(\mu_i)\}^2} + \mathbf{D}^{-1} + \mathbf{R}, \quad (3.30)$$

com

$$\mathbf{R} = - \sum_{i=1}^n (y_i - \mu_i) z_i \frac{\partial}{\partial \mathbf{b}} \left[\frac{1}{\phi a_i v(\mu_i) g'(\mu_i)} \right], \quad (3.31)$$

tendo $\mathbf{R}$, valor esperado igual a zero. As derivações matemáticas de $q'(\mathbf{b})$ e $q''(\mathbf{b})$ encontram-se no apêndice deste capítulo (Seção 3.4).

Ademais, Breslow e Clayton (1993) utilizam, ainda, o conceito de *pesos iterativos* da classe de modelos GLM de McCullagh e Nelder (1989, seção 2.5), a partir do qual estabelecem $w_i = \{\phi a_i v(\mu_i) \{g'(\mu_i)\}^2\}^{-1}$, havendo, para tal, uma matriz diagonal $\mathbf{W}$, $n \times n$, constituída pelos próprios termos w_i . Dessa maneira, $q''(\mathbf{b})$ poderia ser reescrita como

$$q''(\mathbf{b}) \approx \mathbf{Z}' \mathbf{W} \mathbf{Z} + \mathbf{D}^{-1}. \quad (3.32)$$

Ainda com relação ao termo $\mathbf{R}$, Breslow e Clayton (1993) esclarecem que a suposição de valor esperado $\mathbf{0}$ para $\mathbf{R}$ é válida para funções de ligação canônicas conforme descrição de McCullagh e Nelder (1989, p. 32).

No entanto, para um melhor entendimento sobre esse tipo específico de função de ligação, bem como das funções de ligação, em geral, uma breve análise encontra-se no apêndice deste capítulo (Seção 3.4).

Ao multiplicar a expressão (3.32) por $\mathbf{D}$, obtemos $q''(\mathbf{b}) \approx \mathbf{I} + \mathbf{Z}' \mathbf{W} \mathbf{Z} \mathbf{D}$ e, ao substituí-la na função de quase-verossimilhança (3.26), obtemos uma outra aproximação para a expressão $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$, a saber,

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\mathbf{I} + \mathbf{Z}' \mathbf{W} \mathbf{Z} \mathbf{D}| - \frac{1}{2} \left(\sum_{i=1}^n \tilde{d}_i + \tilde{\mathbf{b}}' \mathbf{D}^{-1} \tilde{\mathbf{b}} \right). \quad (3.33)$$

sendo que $\tilde{\mathbf{b}}$, ao minimizar $q(\tilde{\mathbf{b}})$, maximiza $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$ e $\tilde{d}_i$ é d_i com $\mathbf{b}$ sendo substituído por $\tilde{\mathbf{b}}$ via $\mu_i = g^{-1}(x_i \boldsymbol{\beta} + z_i \mathbf{b})$.

De acordo ainda com Breslow e Clayton (1993), se os *pesos iterativos* $w_i = \{\phi a_i v(\mu_i) \{g'(\mu_i)\}^2\}^{-1}$, variam pouco como função da média, o primeiro termo de $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$ pode ser ignorado, de modo que sua expressão pode ser novamente aproximada pela expressão

$$ql_p \approx -\frac{1}{2} \left(\sum_{i=1}^n \tilde{d}_i + \tilde{\mathbf{b}}' \mathbf{D}^{-1} \tilde{\mathbf{b}} \right), \quad (3.34)$$

também conhecida como *penalized quasi-log-likelihood* (PQL).

Jang (2007, p. 130) realiza uma interpretação mais intuitiva acerca da função PQL. A minimização de $q(\mathbf{b})$ ocorre em função da substituição de $\tilde{\mathbf{b}}$ em $q(\mathbf{b})$. Além disso, para dado $\boldsymbol{\beta}$, $\tilde{\mathbf{b}}$ maximiza o log da função de quase-verossimilhança penalizada, ql_p , ainda que $\tilde{\mathbf{b}}$ seja dependente de $\boldsymbol{\beta}$, fato que permite, de certa maneira, escrever $\tilde{\mathbf{b}} = \tilde{\mathbf{b}}(\boldsymbol{\beta})$.

Em contrapartida, para dado θ , se considerarmos $\hat{\beta}$ como o elemento maximizador de ql_p (como função de β), então, temos que $\hat{\beta}$ e $\hat{b}$ maximizam conjuntamente

$$ql_p(\beta, b) = -\frac{1}{2} \left(\sum_{i=1}^n d_i + b' D^{-1} b \right), \quad (3.35)$$

com $\hat{b} = \tilde{b}(\hat{\beta})$.

Breslow e Clayton (1993) reconhecem e atribuem manifestamente que o resultado a que chegaram, dado pela expressão (3.34), já constituía um resultado anteriormente alcançado em Green (1987).

Apenas a título de informação, Green (1987) trabalhou na obtenção de algoritmos capazes de maximizar uma função de verossimilhança penalizada (e não de quase-verossimilhança penalizada), sendo sua expressão dada por $L(\theta(\beta, \gamma)) - \frac{1}{2} \lambda \xi^T K \xi$, em um contexto geral de modelos de regressão.

Por fim, após a elucidação do método que permite conhecer a forma da função PQL, os estimadores para β e b podem ser obtidos através da resolução do sistema de equações não lineares via $\frac{\partial ql_p(\beta, b)}{\partial \beta} = 0$ e $\frac{\partial ql_p(\beta, b)}{\partial b} = 0$, dadas, respectivamente, por

$$\sum_{i=1}^n \frac{(y_i - \mu_i) x_i}{\phi a_i v(\mu_i) g'(\mu_i)} = 0 \quad (3.36)$$

e

$$\sum_{i=1}^n \frac{(y_i - \mu_i) z_i}{\phi a_i v(\mu_i) g'(\mu_i)} - D^{-1} b = 0. \quad (3.37)$$

3.3.3 Algoritmo de Estimação

Para a solução do sistema de equações não lineares (3.36) e (3.37), Breslow e Clayton (1993) propuseram um algoritmo iterativo de escore de Fisher baseado no sistema

$$\begin{pmatrix} X^t W X & X^t W Z D \\ Z^t W X & I + Z^t W Z D \end{pmatrix} \begin{pmatrix} \beta \\ v \end{pmatrix} = \begin{pmatrix} X^t W Y \\ Z^t W Y \end{pmatrix}, \quad (3.38)$$

com $b = Dv$.

Segundo Breslow e Clayton (1993), o procedimento (3.38) foi desenvolvido em Harville (1977) no intuito de obter estimadores BLUE (*best linear unbiased estimation*) para β e b com base na teoria normal multivariada, isto é, em um contexto no qual $Y = X\beta + Zb + \varepsilon$, com $\varepsilon \sim N(0, W^{-1})$ e $b \sim N(0, D)$. De modo equivalente, (3.38) também pode ser solucionado, inicialmente por β , por

$$(X^t V^{-1} X) \boldsymbol{\beta} = X^t V^{-1} Y, \quad (3.39)$$

sendo $V = W^{-1} + Z D Z^t$ e, em seguida,

$$\mathbf{b} = D \mathbf{v} = D Z^t V^{-1} (Y - X \boldsymbol{\beta}). \quad (3.40)$$

Jang (2007, p. 131) chama, ainda, atenção para o fato de que (3.38) corresponde, também, ao sistema de equações dos modelos mistos de Henderson (1950), por meio do qual procurava-se obter estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ e $\mathbf{b}$, tratando-os como fixados, ao diferenciar o logaritmo da função densidade conjunta com relação a $\boldsymbol{\beta}$ e $\mathbf{b}$ e igualando as derivadas parciais a zero. As soluções contidas em Henderson (1950) são as mesmas de (3.39) e (3.40).

3.3.4 Componentes de Variância

Note que, ao longo da apresentação do procedimento de estimação PQL (seção anterior), os componentes de variância, $\boldsymbol{\theta}$, ou, em outros termos, o vetor dos parâmetros de dispersão, ainda que desconhecidos, foram tratados como fixados.

O lado direito das equações (3.39) e (3.40), por exemplo, depende de $\boldsymbol{\theta}$, no caso, mais especificamente de D . Isso significa, tecnicamente, que os elementos $\boldsymbol{\beta}$ e $\mathbf{b}$ das equações (3.43) e (3.44) não podem ser considerados como estimadores, uma vez que dependem de parâmetro desconhecido. Tecnicamente, um estimador é uma função que só pode depender de valores amostrais. Logo, obter expressões para estimar parâmetros tais como $\boldsymbol{\beta}$ e $\mathbf{b}$, sem que esses dependam de outro(s) parâmetro(s), é fundamental, e métodos específicos de estimação de componentes de variância também devem receber atenção.

Os métodos de estimação dos componentes de variância exercem um papel central não só na classe dos modelos GLMM mas, como veremos no capítulo seguinte, na classe dos modelos hierárquicos (HLM).

No caso da classe de modelos GLMM, o método dos Componentes de Variância é essencial para estimar a variância dos efeitos aleatórios, enquanto que, no caso dos modelos hierárquicos, esse método permite estimar a variância dos erros em níveis hierárquicos mais elevados.

Especificamente no caso da classe dos modelos GLMM (ao menos em seu surgimento), a estimação de componentes de variância leva em consideração um tipo de função específica para auxiliar na estimação de parâmetros desconhecidos, a saber, a função de verossimilhança perfilada (*profile likelihood*).

Esse tipo de função pode ser útil em situações nas quais um grande número de parâmetros precisa ser estimado. Isso porque, segundo Pawitan (2001): *“Even when we are interested in several parameters, it is always easier to describe one parameter at a time. The problem also arises in cases where we may be interested in only a subset of the parameters.”* (p. 61)

Para ilustrar melhor, vamos considerar um vetor de parâmetros desconhecidos θ , tal que $\theta = (\theta_1, \theta_2)$, e uma função de verossimilhança $L(\theta_1, \theta_2)$. Logo, se o interesse reside na estimação de, por exemplo, θ_1 , Pawitan (2001, p. 62) chama a atenção para o fato de que se θ_2 puder ser escrito como função de θ_1 , ou seja, $\theta_2(\theta_1)$, a função de verossimilhança perfilada de θ_1 pode ser escrita como

$$L(\theta_1) = L(\theta_1, \hat{\theta}_2(\theta_1)), \quad (3.41)$$

sendo $\hat{\theta}_2(\theta_1)$ a estimativa de máxima verossimilhança de θ_2 para um valor fixo de θ_1 .

Andrade (2013, p. 21) acrescenta, ainda, que a função de verossimilhança perfilada constitui uma espécie de adaptação da própria função de verossimilhança, tendo certos parâmetros substituídos por suas estimativas de máxima verossimilhança.

No caso da classe de modelos GLMM, Breslow e Clayton (1993) aproximam uma função de quase-verossimilhança perfilada para θ , com interesse específico em estimar os componentes de θ (variância de $\mathbf{b}$), escrevendo β como $\beta(\theta)$, de modo que a função de verossimilhança perfilada seja $ql(\theta, \hat{\beta}(\theta))$.

Ao assumir algumas suposições, tais como ignorar a dependência de $\mathbf{W}$ em θ e substituir a função de quase-desvio $\sum d_i(y_i, \mu_i)$ pela estatística de Pearson $\sum (y_i - \mu_i)^2 / (a_i v(\mu_i))$, a suposição, por parte de Breslow e Clayton (1993), de distribuição normal multivariada (para $\mathbf{y}$) feita à seção anterior (Seção 3.3.3) continua válida nesta seção, de forma que a função $ql(\theta, \hat{\beta}(\theta))$ assume a forma

$$ql(\theta, \hat{\beta}(\theta)) \approx -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\hat{\beta})^t \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta}), \quad (3.42)$$

encontrando semelhança com a expressão descrita em Harville (1977, Seção 4), para estimação de θ , em um contexto de função de verossimilhança.

Em seguida, ao levar em consideração os ajustes necessários para obtenção de estimadores de verossimilhança restrita, REML, ao menos na versão específica de Patterson e Thompson (1971), os autores obtêm uma nova forma, ql_1 , a partir de ql , dada por

$$\begin{aligned}
ql_1(\boldsymbol{\theta}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})) &\approx -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}| \\
&\quad - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned} \tag{3.43}$$

encontrando, também, semelhança com a expressão descrita em Harville (1977, Seção 5), para estimação de $\boldsymbol{\theta}$, com função de verossimilhança.

A passagem da forma da expressão de ql para a forma da expressão de ql_1 , isto é, de (3.42) para (3.43), pode ser encontrada no apêndice deste capítulo (Seção 3.4), baseado em dois exemplos de Pawitan (2001).

Em seguida, Breslow e Clayton (1993) observam que as suposições que são necessárias para o ajuste em ql_1 , a saber, que: i) os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ devem ser ortogonais e ii) que a matriz de informação para $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ deve ser $\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}$, não foram satisfeitas para a classe de modelos GLMM, ao menos por essa metodologia desenvolvida pelos próprios autores.

Em decorrência desse fato, Breslow e Clayton (1993) adotam um procedimento descrito em Harville (1977), de forma que as estimativas de máxima verossimilhança restrita (REML) dos componentes de variância para a classe de modelos GLMM passam a ser obtidas ao diferenciar (3.42) e igualá-la a zero, isto é,

$$-\frac{1}{2}\left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \text{tr}\left(\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_i}\right)\right] = 0, \tag{3.44}$$

sendo que $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}$.

Para a inferência de $\boldsymbol{\theta}$, a matriz de informação de Fisher assume a forma

$$\mathbb{I}_{jk} = -\frac{1}{2}\text{tr}\left(\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_k}\right). \tag{3.45}$$

Em relação à classe dos modelos hierárquicos (assunto do próximo capítulo), o método de estimação para os componentes de variância foi implementado (pelos seus proponentes) de modo distinto, assunto também a ser analisado.

3.4 Apêndice

- Demonstrações (3.2) e (3.3)

Os $\mathbf{b}_i$ são distribuídos segundo a distribuição normal com média $\mathbf{0}$ e matriz de covariância positiva definida $\mathbf{D}$, de dimensão $k \times k$, isto é, $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$. Os erros $\mathbf{e}_i$'s são independentes,

distribuídos segundo a distribuição normal com média $\mathbf{0}$ e matriz de covariância $\mathbf{R}_i$ positiva definida, com dimensão $n_i \times n_i$, isto é, $\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. Dessa maneira, temos que

$$\begin{aligned}\mathbb{E}(\mathbf{y}_i) &= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\ &= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta}) + \mathbb{E}(\mathbf{Z}_i\mathbf{b}_i) + \mathbb{E}(\mathbf{e}_i) \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbb{E}(\mathbf{b}_i) + \mathbf{0} \\ &= \mathbf{X}_i\boldsymbol{\beta}.\end{aligned}$$

$$\begin{aligned}\text{Var}(\mathbf{y}_i) &= \text{Var}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\ &= \text{Var}(\mathbf{X}_i\boldsymbol{\beta}) + \text{Var}(\mathbf{Z}_i\mathbf{b}_i) + \text{Var}(\mathbf{e}_i) \\ &= \mathbf{Z}_i\text{Var}(\mathbf{b}_i)\mathbf{Z}_i^t + \mathbf{R}_i \\ &= \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^t + \mathbf{R}_i.\end{aligned}$$

- Demonstração de (3.14)

O primeiro estágio, pela perspectiva bayesiana no intuito de obter estimativas $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R)$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}_R)$, permanece praticamente inalterado, com a observação de que, condicionado em $\boldsymbol{\beta}$ e $\mathbf{b}_i$, a variável resposta $\mathbf{y}_i$ possui distribuição normal com média $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$ e matriz de variância $\mathbf{R}_i$ ou seja,

$$\begin{aligned}\mathbb{E}(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{b}_i) &= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\ &= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta}) + \mathbb{E}(\mathbf{Z}_i\mathbf{b}_i) + \mathbb{E}(\mathbf{e}_i) \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i.\end{aligned}$$

$$\begin{aligned}\text{Var}(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{b}_i) &= \text{Var}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\ &= \text{Var}(\mathbf{X}_i\boldsymbol{\beta}) + \text{Var}(\mathbf{Z}_i\mathbf{b}_i) + \text{Var}(\mathbf{e}_i) \\ &= \mathbf{R}_i.\end{aligned}$$

- Demonstração de (3.15)

Já no segundo estágio, pela perspectiva bayesiana no intuito de obter estimativas $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_R)$ e $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}_R)$, os parâmetros $\boldsymbol{\beta}$ e $\mathbf{b}_i$ são variáveis aleatórias independentes, normalmente distribuídas, ambas com média $\mathbf{0}$ e matrizes de variância $\text{Var}(\boldsymbol{\beta}) = \boldsymbol{\Gamma}$ e $\text{Var}(\mathbf{b}_i) = \mathbf{D}$, sendo que a independência entre $\boldsymbol{\beta}$ e $\mathbf{b}_i$ implica $\text{Cov}(\boldsymbol{\beta}, \mathbf{b}_i) = \mathbf{0}$.

$$\begin{aligned}
\mathbb{E}(\mathbf{y}_i) &= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\
&= \mathbb{E}(\mathbf{X}_i\boldsymbol{\beta}) + \mathbb{E}(\mathbf{Z}_i\mathbf{b}_i) + \mathbb{E}(\mathbf{e}_i) \\
&= \mathbf{X}_i\mathbb{E}(\boldsymbol{\beta}) + \mathbf{Z}_i\mathbb{E}(\mathbf{b}_i) + \mathbb{E}(\mathbf{e}_i) \\
&= \mathbf{0}.
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\mathbf{y}_i) &= \text{Var}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i) \\
&= \text{Var}(\mathbf{X}_i\boldsymbol{\beta}) + \text{Var}(\mathbf{Z}_i\mathbf{b}_i) + \text{Var}(\mathbf{e}_i) \\
&= \mathbf{X}_i\text{Var}(\boldsymbol{\beta})\mathbf{X}_i^t + \mathbf{Z}_i\text{Var}(\mathbf{b}_i)\mathbf{Z}_i^t + \mathbf{R}_i \\
&= \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i^t + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^t + \mathbf{R}_i.
\end{aligned}$$

- A função quase-desvio

A função quase-desvio, d_i , denota uma medida (ou critério) da qualidade do ajuste do modelo denominada *deviance*.

McCulloch e Nelder (1989) estabelecem que o critério *deviance*, mais precisamente *scaled deviance*, é dado por

$$D^*(\mathbf{y}; \mu) = 2l(\mathbf{y}; \mathbf{y}) - 2l(\mu; \mathbf{y}),$$

sendo $l(\mathbf{y}; \mathbf{y})$ o máximo (do logaritmo da função) de verossimilhança alcançável para um modelo em que os valores ajustados são exatamente iguais aos dados observados. Uma vez que $l(\mathbf{y}; \mathbf{y})$ não depende de parâmetros, temos que maximizar $l(\mu; \mathbf{y})$ equivale a minimizar $D^*(\mathbf{y}; \mu)$, no que diz respeito a μ .

Apenas a título de ilustração, para uma única observação y de uma distribuição normal com variância σ^2 conhecida, temos que

$$f(y; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\}.$$

O logaritmo da função de verossimilhança será

$$l(\mu; y) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y - \mu)^2}{2\sigma^2}.$$

Ao assumirmos $\mu = y$, obteremos o máximo do logaritmo da função de verossimilhança, isto é,

$$l(y; y) = -\frac{1}{2}\log(2\pi\sigma^2).$$

Logo, o critério *deviance* será dado por

$$\begin{aligned}
D^*(y; \mu) &= 2l(y; y) - 2l(\mu; y) \\
&= -\log(2\pi\sigma^2) - 2 \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y - \mu)^2}{2\sigma^2} \right] \\
&= \frac{(y - \mu)^2}{\sigma^2}.
\end{aligned}$$

Breslow e Clayton (1993) consideram o critério *scaled deviance* sob a forma $2\phi\{l(y; y, \phi) - l(\mu; y, \phi)\}$. Se considerarmos que o parâmetro de dispersão ϕ é o próprio σ^2 (conhecido), teríamos, então, que o critério *scaled deviance*, para o caso da distribuição normal, seria igual a

$$2\phi\{l(y; y, \phi) - l(\mu; y, \phi)\} = \sigma^2 \left\{ \frac{(y - \mu)^2}{\sigma^2} \right\} = (y - \mu)^2,$$

o que, conforme constatado, coincidiria com a soma do quadrado dos resíduos, a não ser por estarmos considerando, especificamente neste momento, o caso de apenas uma única observação.

- Derivada primeira de $q(\mathbf{b})$

Na expressão (3.25), isto é,

$$q(\mathbf{b}) = \frac{1}{2} \left(\sum_{i=1}^n d_i + \mathbf{b}^t \mathbf{D}^{-1} \mathbf{b} \right),$$

a derivada do segundo termo (dentro dos parênteses) produz $2\mathbf{D}^{-1}\mathbf{b}$ que, multiplicado por $\frac{1}{2}$ (constante fora dos parênteses), resulta em $\mathbf{D}^{-1}\mathbf{b}$.

A seguir, serão realizados os passos do processo de derivação do primeiro termo dentro dos parênteses ($\sum d_i$).

Lembrando que $d_i(y; \mu) = -2 \int_y^\mu \frac{y-u}{\phi a_i v(u)} du$, a simples multiplicação de -2 de d_i também por $\frac{1}{2}$ (constante fora dos parênteses) resultará em um somatório (de integrais) com sinal negativo (-1).

Uma vez que $q(\mathbf{b})$ envolve a soma de todos os d_i , ou seja, $\sum_{i=1}^n d_i$ e, como a derivada de uma soma é a soma de derivadas, temos que

$$\frac{\partial}{\partial \mathbf{b}} \sum_{i=1}^n d_i = \sum_{i=1}^n \frac{\partial d_i}{\partial \mathbf{b}}.$$

Ao substituir d_i pela expressão da integral e levar em consideração seu sinal negativo, a soma das derivadas pode ser reescrita como

$$\sum_{i=1}^n \frac{\partial d_i}{\partial \mathbf{b}} = - \sum_{i=1}^n \frac{\partial}{\partial \mathbf{b}} \left(\int_{y_i}^{\mu_i} \frac{y_i - t}{\phi a_i v(u)} dt \right).$$

Ao denominar a integral

$$F(\mu_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi a_i v(u)} dt,$$

e sem perder de vista que, com relação ao valor esperado da variável resposta, μ_i , a função de ligação $g(\cdot)$, avaliada em μ_i , estabelece que $g(\mu_i) = x_i \boldsymbol{\beta} + z_i \mathbf{b}$, logo, por meio da função inversa $h = g^{-1}$, pode-se especificar $\mu_i = h(x_i \boldsymbol{\beta} + z_i \mathbf{b})$.

Assim, uma vez que a derivada para a expressão de d_i implica na derivada de uma integral, temos que a derivada dessa integral, $F(\mu_i)$, passa a ser dada por $\frac{dF}{d\mathbf{b}} = \frac{dF}{d\mu_i} \frac{d\mu_i}{d\mathbf{b}}$, de modo que

$$\frac{dF}{d\mu_i} = \frac{d}{d\mu_i} \left[\int_{y_i}^{\mu_i} \frac{y_i - t}{\phi a_i v(u)} dt \right].$$

Por meio do segundo teorema fundamental do cálculo, temos que

$$\frac{d}{d\mu_i} \left[\int_{y_i}^{\mu_i} \frac{y_i - t}{\phi a_i v(u)} dt \right] = \frac{y_i - \mu_i}{\phi a_i v(u)}. \quad (3.50)$$

Já com relação à derivada $\frac{d\mu_i}{d\mathbf{b}}$ e, lembrando, ainda, que $\mu_i = h(x_i \boldsymbol{\beta} + z_i \mathbf{b})$,

$$\frac{d\mu_i}{d\mathbf{b}} = \frac{d\mu_i}{dh} \frac{dh}{d\mathbf{b}} = h'(x_i \boldsymbol{\beta} + z_i \mathbf{b}) z_i.$$

Entretanto, uma vez que $h = g^{-1}$, temos que a derivada de h , isto é, h' , equivale, na prática, à derivada de uma função inversa $(g^{-1})'$ que, por regra, é obtida pela inversa de sua derivada, no domínio de g . Isso significa que $h'(x_i \boldsymbol{\beta} + z_i \mathbf{b}) = (g'(\mu_i))^{-1} = \frac{1}{g'(\mu_i)}$.

Desse modo, teremos

$$\frac{d\mu_i}{d\mathbf{b}} = \frac{d\mu_i}{dh} \frac{dh}{d\mathbf{b}} = \frac{1}{g'(\mu_i)} z_i. \quad (3.51)$$

Logo, ao combinar (3.45) e (3.46), obtemos

$$\frac{dF}{d\mathbf{b}} = \frac{y_i - \mu_i}{\phi a_i v(\mu_i)} \frac{z_i}{g'(\mu_i)},$$

de forma que $q'(\mathbf{b}) = 0$ se, e somente se,

$$\mathbf{D}^{-1} \mathbf{b} - \sum_{i=1}^n \frac{(y_i - \mu_i)}{\phi a_i v(\mu_i)} \frac{z_i}{g'(\mu_i)} = 0.$$

- Derivada segunda de $q(\mathbf{b})$

Para obter a derivada segunda, $q''(\mathbf{b})$, é preciso diferenciar $q'(\mathbf{b})$ mais uma vez em relação a $\mathbf{b}$, ou seja,

$$q''(\mathbf{b}) = - \sum_{i=1}^n \frac{\frac{1}{g'(\mu_i)} z_i z_i [\phi a_i v(\mu_i) g'(\mu_i)] - (y_i - \mu_i) z_i \frac{\partial}{\partial \mathbf{b}} [\phi a_i v(\mu_i) g'(\mu_i)]}{[\phi a_i v(\mu_i) g'(\mu_i)]^2} + \mathbf{D}^{-1}.$$

Ao repartir em dois somatórios, teremos

$$q''(\mathbf{b}) = \sum_{i=1}^n \frac{\frac{1}{g'(\mu_i)} z_i z_i [\phi a_i v(\mu_i) g'(\mu_i)]}{[\phi a_i v(\mu_i) g'(\mu_i)]^2} - \sum_{i=1}^n \frac{(y_i - \mu_i) z_i \frac{\partial}{\partial \mathbf{b}} [\phi a_i v(\mu_i) g'(\mu_i)]}{[\phi a_i v(\mu_i) g'(\mu_i)]^2} + \mathbf{D}^{-1}.$$

Simplificando alguns termos, obtemos

$$q''(\mathbf{b}) = \sum_{i=1}^n \frac{\frac{1}{g'(\mu_i)} z_i z_i}{\phi a_i v(\mu_i) g'(\mu_i)} - \sum_{i=1}^n (y_i - \mu_i) z_i \frac{\partial}{\partial \mathbf{b}} \left[\frac{1}{\phi a_i v(\mu_i) g'(\mu_i)} \right] + \mathbf{D}^{-1},$$

de modo que

$$q''(\mathbf{b}) = \sum_{i=1}^n \frac{z_i z_i}{\phi a_i v(\mu_i) [g'(\mu_i)]^2} - \sum_{i=1}^n (y_i - \mu_i) z_i \frac{\partial}{\partial \mathbf{b}} \left[\frac{1}{\phi a_i v(\mu_i) g'(\mu_i)} \right] + \mathbf{D}^{-1},$$

tendo Breslow e Clayton (1993) denominado o segundo termo do lado direito da igualdade como $\mathbf{r}$. Logo,

$$q''(\mathbf{b}) = \sum_{i=1}^n \frac{z_i z_i}{\phi a_i v(\mu_i) [g'(\mu_i)]^2} - \mathbf{R} + \mathbf{D}^{-1}.$$

Com a suposição do valor esperado de $\mathbf{R}$ igual a zero (em probabilidade e na medida em que n cresce), a derivada segunda pode ser aproximada por

$$q''(\mathbf{b}) \approx \sum_{i=1}^n \frac{z_i z_i}{\phi a_i v(\mu_i) [g'(\mu_i)]^2} + \mathbf{D}^{-1}.$$

- Funções de ligação canônicas

Inicialmente seria interessante destacar o papel que as funções de ligação, de um modo geral, exercem na classe dos modelos GLM.

McCullagh e Nelder (1989, p. 31) atentam para o fato de que, na classe dos modelos (clássicos) lineares (LM), a média μ_i e o preditor linear η_i são iguais, uma vez que podem

assumir valores reais. Logo, a função de ligação *identidade* é, por si só, capaz de lidar com classe de modelos de modo satisfatório.

Contudo, ao se defrontar, por exemplo, com dados de contagem (e suposição de distribuição de Poisson para a variável resposta), μ_i deve, necessariamente, assumir valores positivos, tornando a função de ligação *identidade* particularmente indesejável. Assim, uma função de ligação deve ser capaz de associar, de forma apropriada, o preditor linear com a média da distribuição da variável resposta.

Note que, se a distribuição da variável resposta pertence à família exponencial, então, sua função densidade poderá ser escrita como

$$f(y_i|\theta_i; \phi) = \exp \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\}.$$

Nesse caso, temos que $b'(\theta_i) = \mu_i = \mathbb{E}(Y_i)$. Além disso, em uma classe de modelos GLM (apenas por uma questão de simplicidade e didatismo), teremos

$$g(\mu_i) = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}.$$

Contudo, se $b(\theta_i) = \theta_i$, o parâmetro desconhecido θ será dito *canônico*, de forma que, nesse caso em particular,

$$\theta_i = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}.$$

Nesse caso, η_i será considerada a *função de ligação canônica* (Paula (2013, p. 7)). Alguns exemplos conhecidos de funções de ligação canônicas são:

- i) distribuição normal, $\eta = \mu$ (função de ligação *identidade*);
- ii) distribuição de Poisson, $\eta = \log \mu$;
- iii) distribuição binomial, $\eta = \log \left\{ \frac{\mu}{1-\mu} \right\}$;
- iv) distribuição gama, $\eta = \mu^{-1}$;
- v) distribuição normal inversa, $\eta = \mu^{-2}$.

- Função de quase-verossimilhança perfilada para obtenção de estimadores REML

Pawitan (2001, p. 440) apresenta uma situação semelhante à expressão (3.42), baseada em um modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e},$$

no qual $\mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{b} \sim N(\mathbf{0}, \mathbf{D})$, com independência entre $\mathbf{b}$ e $\mathbf{e}$.

A distribuição marginal de $\mathbf{y}$, também com distribuição normal, possui média $\mathbf{X}\boldsymbol{\beta}$ e variância $\mathbf{V} = \boldsymbol{\Sigma} + \mathbf{ZDZ}^t$, com passagens parecidas com aquelas encontradas em (3.2) e (3.3). Logo, a função log-verossimilhança para os parâmetros $(\boldsymbol{\beta}, \boldsymbol{\theta})$ é dada por

$$\log L(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

A derivada de $\log L(\boldsymbol{\beta}, \boldsymbol{\theta})$ com relação a $\boldsymbol{\beta}$ é dada por

$$(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}) \boldsymbol{\beta} = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y},$$

resultando na expressão dos mínimos quadrados ponderados (ou generalizados) e, ainda, em conformidade com a primeira fórmula do sistema de equações (3.37).

Se considerarmos, ainda, o exemplo 10.12 do mesmo autor, na página 291, é possível considerar

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}, \quad (3.52)$$

de modo que, ao adaptar para uma função de verossimilhança perfilada, especificamente para o parâmetro $\boldsymbol{\theta}$, teremos

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}})^t \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}).$$

com Informação de Fisher $I(\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}$.

Apesar de não haver uma fórmula geral para $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}_{\boldsymbol{\theta}}}$ (operação a ser realizada no processo de obtenção da derivada segunda), é possível verificar que

$$\mathbb{E} \left(\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}_i} \log L(\boldsymbol{\beta}, \boldsymbol{\theta}) \right) = \mathbb{E} \left(\mathbf{X}^t \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) = 0,$$

para qualquer $\boldsymbol{\theta}_i$, de modo que $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ são informações ortogonais e o jacobiano $\left| \frac{\partial \hat{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}_{\boldsymbol{\theta}}} \right| \approx 1$.

Logo, a função de verossimilhança perfilada modificada será dada por

$$\begin{aligned} \log L_m(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) - \frac{1}{2} \log |\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}| \\ &= -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}})^t \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}), \end{aligned}$$

coincidindo exatamente com a expressão obtida por Patterson e Thompson (1971) para obter estimativas REML.

4 O MODELO LINEAR HIERÁRQUICO GENERALIZADO (HGLM)

4.1 Introdução

Nesta seção, será introduzida a classe de Modelos Lineares Hierárquicos Generalizados (*Hierarchical Generalized Linear Models* – HGLM). Mas, inicialmente, será feita uma introdução à classe de modelos lineares hierárquicos (*Hierarchical Linear Models* – HLM). Essa classe de modelos (HLM) também representa uma extensão da classe dos modelos tradicionais de regressão linear, uma vez que leva em consideração e, portanto, é capaz de acomodar, uma estrutura de dados com mais de um nível de organização (hierárquica). Os métodos de estimação dessas classes de modelos serão discutidos. Por fim, uma análise acerca da relação entre as classes de modelos HGLM e GLMM também será realizada.

4.2 O modelo HLM

De acordo com Gelman e Hill (2007, p. 2), os modelos Hierárquicos (ou Multiníveis) são assim chamados por dois motivos:

- Pela estrutura dos dados; e
- Pelo próprio modelo, uma vez que possui uma hierarquia organizada.

Esses modelos frequentemente recebem outras denominações, a depender da área em que estão sendo empregados. Na Sociologia, por exemplo, são comumente denominados *Modelos Lineares Multiníveis*. Na Biometria, são designados como *Modelos de Efeitos Mistos* e ou como *Modelos de Coeficientes Aleatórios*. Na Econometria, são referidos como Modelos de Regressão com Coeficientes Aleatórios, enquanto que, na literatura estatística, são conhecidos como Modelos de Componentes de Covariância (Raudenbush e Bryk (2002, p. 6)).

Segundo, ainda, Raudenbush e Bryk (2002), a expressão Modelos Lineares Hierárquicos foi alcunhada por Lindley e Smith (1972) e Smith (1973) em razão de suas relevantes contribuições acerca de processos de estimação de modelos lineares, via métodos bayesianos.

Lindley e Smith (1972) propuseram uma estrutura de modelo bayesiano, baseado no método dos mínimos quadrados, a partir do qual os próprios parâmetros do modelo possuíam uma estrutura linear geral, em termos de hiperparâmetros. Ainda nesse mesmo artigo, os autores, na Seção 2, deram um passo adiante no sentido de permitirem que os hiperparâmetros

também admitissem uma estrutura linear geral, resultando, de acordo com as próprias palavras dos autores, em um modelo de três estágios.

Smith (1973) descobre e analisa certas propriedades que são características dos estimadores do modelo de Lindley e Smith (1972), bem como aprimora e estende os métodos de estimação envolvidos nos estágios, vale dizer, nos níveis hierárquicos.

Entretanto, de acordo com Woltman, Feldstain, MacKay e Rocchi (2012, p. 52), foi a partir de Dempster, Laird e Rubin (1977) que se tornaram possíveis a disseminação e a efetiva utilização dos modelos lineares hierárquicos, uma vez que esses últimos autores ofereceram uma solução para a estimação de componentes de variância, especificamente na presença de dados desbalanceados, característica dessa classe de modelos.

Assim, se por um lado, autores como Lindley e Smith (1972) e Smith (1973) desenvolveram uma nova classe de modelos (os modelos lineares hierárquicos), por outro, Dempster et al. (1977) ofereceram uma solução para o problema de estimação de componentes de variância via algoritmo E-M (já descrito anteriormente no Capítulo 3, durante a descrição dos modelos LMM). Em essência, a estimação dos componentes de variância na classe dos modelos lineares hierárquicos foi solucionada, inicialmente, por meio de algoritmos E-M, ao considerar os próprios componentes de variância (variância dos erros nos distintos níveis hierárquicos) como dados faltantes (*missing data*).

A seguir, será descrita, com mais detalhes, a lógica interna dessa classe de modelo.

4.2.1 O Modelo HLM em Dois Níveis

Por razões de simplicidade e didatismo, será apresentado um modelo hierárquico simples, isto é, com dois níveis hierárquicos, tal como em Raudenbush e Bryk (2002), Capítulo 3. Esse modelo pode perfeitamente ser estendido para um modelo com grau maior de complexidade, isto é, de níveis hierárquicos mais elevados (3 ou 4 níveis, por exemplo).

Raudenbush e Bryk (2002) fazem uma primeira introdução mais elementar, isto é, de uma perspectiva da ANOVA de um Fator. Em seguida, estendem o modelo para um nível hierárquico superior (Nível 2). Os métodos de estimação em ambos os níveis também serão considerados. Com relação a notação, o termo Y_{ij} refere-se à variável resposta da i -ésima UE (Nível 1), pertencente ao j -ésimo nível hierárquico (Nível 2) de organização.

Uma versão simplificada para um modelo Nível 1 pode ser dada por

$$Y_{ij} = \beta_{0j} + r_{ij}, \quad (4.1)$$

com o intercepto β_{0j} podendo ser interpretado como a resposta esperada (média) de uma unidade individual (de Nível-1) pertencente a uma estrutura hierárquica mais elevada (de Nível 2), com $r_{ij} \sim N(0, \sigma^2)$.

Ao tomarmos uma unidade do Nível 2 (por exemplo, a j -ésima), teremos, em termos médios,

$$\bar{Y}_j = \beta_{0j} + \bar{r}_j, \quad (4.2)$$

sendo $\bar{Y}_j$ a média das respostas observadas das unidades individuais (Nível 1) contidas na j -ésima unidade de Nível 2 e $\bar{r}_j$ representando o erro médio (de estimação), com $\bar{r}_j = \sum_{i=1}^{n_j} \frac{r_{ij}}{n_j}$, com variância dada por

$$Var(\bar{r}_j) = \frac{\sigma^2}{n_j} = V_j. \quad (4.3)$$

É importante notar que, nesse contexto, a própria média amostral $\bar{Y}_j$ constitui uma estimativa para β_{0j} .

Uma extensão natural para um modelo Nível 2 consiste na suposição de efeitos sobre o intercepto, vale dizer, das médias das respostas no Nível 1, provenientes de níveis hierárquicos mais elevados. Nesse caso, um possível modelo para o Nível 2 seria

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (4.4)$$

com γ_{00} representando a média global (da população) e u_{0j} descrevendo um efeito aleatório associado com a j -ésima unidade, com $u_{0j} \sim N(0, \tau_{00})$.

Ao combinar (4.4) e (4.2), teremos o modelo

$$\bar{Y}_j = \gamma_{00} + u_{0j} + \bar{r}_j, \quad (4.5)$$

com $\bar{Y}_j$ sendo explicitamente descrito, no caso, por uma média global (populacional).

Devemos notar, no caso, que a variância de $\bar{Y}_j$ conterá dois componentes de variância: i) variância do parâmetro, τ_{00} , e ii) variância do erro, V_j , isto é,

$$\begin{aligned} Var(\bar{Y}_j) &= Var(u_{0j}) + Var(\bar{r}_j) \\ &= \tau_{00} + V_j \\ &= \Delta_j. \end{aligned} \quad (4.6)$$

Raudenbush e Bryk (2002) chamam atenção especial para o fato de que a variância do erro, $Var(\bar{r}_j) = \frac{\sigma^2}{n_j} = V_j$, pode variar de acordo com o tamanho amostral n_j de cada unidade do Nível 2.

Se todas as unidades do Nível 2 tiverem a mesma quantidade de observações, isto é, se $n_j = n$, para todo $j = 1, \dots, J$, então, teremos $Var(\bar{r}_j) = \frac{\sigma^2}{n} = V$, de modo que a variância de ambos os componentes será igual, $\Delta = \tau_{00} + V$, em todas as unidades do Nível 2. Nesse caso em particular, Raudenbush e Bryk (2002) enunciam que o único estimador não viesado de variância mínima para a média global γ_{00} será simplesmente a média aritmética de todos os valores de $\bar{Y}_j$, ou seja,

$$\hat{\gamma}_{00} = \frac{\sum_{j=1}^J \bar{Y}_j}{J}. \quad (4.7)$$

Do contrário, isto é, se não houver igualdade entre os n_j 's, o estimador de γ_{00} será ponderado por uma medida de precisão de $\bar{Y}_j$ dada por

$$Precisão(\bar{Y}_j) = \Delta_j^{-1}, \quad (4.8)$$

e passa a assumir a forma

$$\hat{\gamma}_{00} = \frac{\sum_{j=1}^J \Delta_j^{-1} \bar{Y}_j}{\sum_{j=1}^J \Delta_j^{-1}}, \quad (4.9)$$

comumente conhecido como estimador de *mínimos quadrados ponderados*, o qual guarda equivalência com o método de máxima verossimilhança (Raudenbush e Bryk (2002, p. 40) e Agresti (2015, p. 140)).

A estrutura do modelo Nível 2 também pode ser ampliada ao incorporar variáveis preditoras e não somente modelar o intercepto β_{0j} em termos de γ_{00} e u_{0j} . Assim, ao introduzirmos uma variável preditora W_j , o modelo Nível 2 assume a forma

$$\bar{Y}_j = \gamma_{00} + \gamma_{01}W_j + u_{0j} + \bar{r}_j. \quad (4.10)$$

A variância de $\bar{Y}_j$, dado W_j , será dada por

$$\begin{aligned} Var(\bar{Y}_j) &= \tau_{00} + V_j \\ &= \Delta_j, \end{aligned} \quad (4.11)$$

com a observação de que, nesse caso, Δ_j é variância residual, ou seja, é a variância condicional de $\bar{Y}_j$, dado W_j . As suposições acerca de u_{0j} e $\bar{r}_j$ permanecem as mesmas.

Com relação aos estimadores para γ_{00} e γ_{01} , novamente se faz necessário analisar o tamanho amostral n_j . Se todas as unidades possuem o mesmo tamanho amostral, então, $\Delta_j = \Delta$, para todo $j = 1, \dots, J$, de modo que os estimadores obtidos para γ_{00} e γ_{01} são ambos de *mínimos quadrados*, dados por

$$\tilde{\gamma}_{01} = \frac{\sum_{j=1}^J (W_j - \bar{W}_{\cdot}) (\bar{Y}_j - \bar{Y}_{\cdot})}{\sum_{j=1}^J (W_j - \bar{W}_{\cdot})^2}, \quad (4.12)$$

$$\text{com } \bar{W}_{\cdot} = \sum_{j=1}^J \frac{W_j}{J} \text{ e } \bar{Y}_{\cdot} = \frac{\sum_{j=1}^J \bar{Y}_j}{J}$$

$$\tilde{\gamma}_{00} = \bar{Y}_{\cdot} - \tilde{\gamma}_{01} \bar{W}_{\cdot}. \quad (4.13)$$

Entretanto, se os tamanhos amostrais n_j forem distintos, os estimadores tanto para γ_{00} quanto para γ_{01} serão de *mínimos quadrados ponderados*, dados por

$$\hat{\gamma}_{01} = \frac{\sum_{j=1}^J \Delta_j^{-1} (W_j - \bar{W}_{\cdot}^*) (\bar{Y}_j - \bar{Y}_{\cdot}^*)}{\sum_{j=1}^J \Delta_j^{-1} (W_j - \bar{W}_{\cdot}^*)^2}, \quad (4.14)$$

$$\text{com } \bar{W}_{\cdot}^* = \frac{\sum_{j=1}^J \Delta_j^{-1} W_j}{\sum_{j=1}^J \Delta_j^{-1}} \text{ e } \bar{Y}_{\cdot}^* = \frac{\sum_{j=1}^J \Delta_j^{-1} \bar{Y}_j}{\sum_{j=1}^J \Delta_j^{-1}}$$

$$\hat{\gamma}_{00} = \bar{Y}_{\cdot}^* - \hat{\gamma}_{01} \bar{W}_{\cdot}^*. \quad (4.15)$$

No caso de haver várias variáveis (preditoras) no Nível 1, o processo de estimação para β continua a ser pelo método *dos mínimos quadrados*.

Já no caso de haver mais de uma preditora no Nível 2, o método *dos mínimos quadrados* continua a ser utilizado para a estimação de γ quando os dados forem perfeitamente balanceados, isto é, sempre que $n_j = n$, para todo $j = 1, \dots, J$. Contudo, se os dados forem desbalanceados, a obtenção das estimativas de γ dar-se-á pelo método *dos mínimos quadrados generalizados* (e não pelo método dos *mínimos quadrados ponderados*).

Uma forma de notar a relação entre os métodos de *mínimos quadrados ponderados* e *mínimos quadrados generalizados* diz respeito à matriz de precisão, Δ_j^{-1} . No primeiro caso, essa matriz é diagonal, enquanto que, no segundo caso, ela deve ser simétrica e positiva definida.

4.2.2 Estimação dos Componentes de Variância

Ao longo deste capítulo, os componentes de variância vinham sendo considerados como conhecidos, apenas no intuito de desenvolver a lógica dos HLM.

Nesta seção, será desenvolvida a lógica de central da técnica desenvolvida especificamente em Dempster et al. (1977), de estimar componentes de variância com base em funções de verossimilhança (supondo a possibilidade de atribuir uma distribuição de probabilidades para a variável resposta) via algoritmo E-M, o qual é apropriado para lidar com variáveis latentes.

Dempster et al. (1977) chamam atenção para o fato de estarem se referindo a distribuições pertencentes à família exponencial, ou seja,

$$f(\mathbf{x}|\boldsymbol{\phi}) = b(\mathbf{x})\exp(\boldsymbol{\phi}t(\mathbf{x})^T)/a(\boldsymbol{\phi}), \quad (4.16)$$

de modo que, nessa expressão de família exponencial, em particular, é possível notar que os autores multiplicaram o parâmetro desconhecido, no caso $\boldsymbol{\phi}$, por uma estatística suficiente $t(\mathbf{x})$.

Antes de conduzir a análise, porém, é importante ressaltar que $\mathbf{x}$, além de não ser diretamente observado (no caso, apenas indiretamente por meio de $\mathbf{y}$), representa, segundo os autores, o que tradicionalmente se chama de parâmetros. Assim, especificamente nessa seção do artigo, $\mathbf{x}$ não representa valores observados de uma variável independente, como se assume normalmente na literatura da Estatística, mas, sim, valores de parâmetros. Logo, $f(\mathbf{x}|\boldsymbol{\phi})$ refere-se à distribuição (densidade) de probabilidade de um parâmetro, em função de um outro parâmetro (hiperparâmetro), fato que implica estarmos especificamente no campo da Inferência Bayesiana.

Logo no início do artigo, os autores postularam uma relação entre uma densidade amostral $f(\mathbf{x}|\boldsymbol{\phi})$, caracterizada como *complete-data specification*, e uma outra densidade amostral $g(\mathbf{y}|\boldsymbol{\phi})$, caracterizada como *incomplete-data specification*, dada por

$$g(\mathbf{y}|\boldsymbol{\phi}) = \int f(\mathbf{x}|\boldsymbol{\phi})d\mathbf{x}, \quad (4.17)$$

de modo que o algoritmo E-M tenha por finalidade encontrar o valor de $\boldsymbol{\phi}$ que maximiza $g(\mathbf{y}|\boldsymbol{\phi})$, dado $\mathbf{y}$ observado, fazendo, para isso, uso primordial de uma família associada $f(\mathbf{x}|\boldsymbol{\phi})$.

Isto posto, se (4.16) estiver assegurado, uma breve ilustração do algoritmo E-M pode ser descrita da seguinte maneira: Considere $\boldsymbol{\phi}^{(p)}$ como o valor atual de $\boldsymbol{\phi}$ após p ciclos do algoritmo, podendo o ciclo seguinte ser realizado por meio dos passos:

- E-passo: Estimar a estatística suficiente $t(\mathbf{x})$ por:

$$\mathbf{t}^{(p)} = \mathbb{E}(\mathbf{t}(\mathbf{x})|\mathbf{y}, \boldsymbol{\phi}^{(p)}). \quad (4.18)$$

- M-passo: Determinar $\boldsymbol{\phi}^{(p+1)}$ como solução das equações

$$\mathbb{E}(\mathbf{t}(\mathbf{x})|\boldsymbol{\phi}) = \mathbf{t}^{(p)}, \quad (4.19)$$

sendo $\mathbf{t}^{(p)}$ as estatísticas suficientes referentes aos dados completos, (com $t^{(p)}$ para o caso univariado). Assim, se $\mathbf{t}^{(p)}$ forem calculadas com base em valores observados de $\mathbf{x}$ obtidos a partir de (4.16), então, as equações (4.19) caracterizam um estimador de máxima-verossimilhança para $\boldsymbol{\phi}$.

Na Seção 4.4 do artigo de Dempster et al. (1977), os autores consideraram uma matriz de planejamento $\mathbf{A}$ conhecida $n \times r$ e $\mathbf{y}$ um vetor $n \times 1$, tal que

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (4.20)$$

sendo $\mathbf{x}$ um vetor (não observável) de dimensão $r \times 1$.

Em seguida, assumiram que tanto $\mathbf{A}$ quanto $\mathbf{x}$ poderiam ser particionados, ou seja,

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{k+1}) \quad (4.21)$$

e

$$\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_{k+1}^T)^T, \quad (4.22)$$

tendo $\mathbf{A}_i$ e $\mathbf{x}_i$ dimensões $n \times r_i$ e $r_i \times 1$, respectivamente, para $i = 1, \dots, k+1$ e $\sum_{i=1}^{k+1} r_i = r$.

Para $\mathbf{x}_i$, os autores assumiram que

$$\mathbf{x}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}), \quad i = 1, \dots, k+1, \quad (4.23)$$

tendo σ_i^2 's como parâmetros desconhecidos, e, para $\mathbf{y}$, distribuição normal com média vetor zero e matriz de covariância, isto é,

$$\Sigma = \sigma_1^2 \Sigma_1 + \sigma_2^2 \Sigma_2 + \dots + \sigma_{k+1}^2 \Sigma_{k+1}, \quad (4.24)$$

com $\Sigma_i = \mathbf{A}_i \mathbf{A}_i^T$ fixos e conhecidos.

O objetivo principal consistia em estimar os componentes de variâncias desconhecidos $\sigma_1^2, \sigma_2^2, \dots, \sigma_{k+1}^2$. Suponha, por um lado, que $\mathbf{A}_{k+1} = \mathbf{I}$, de modo que (4.24) seja reescrita sob a forma

$$\mathbf{y} = \sum_{i=1}^k \mathbf{A}_i \mathbf{x}_i + \mathbf{x}_{k+1}. \quad (4.25)$$

Dempster et al. (1977) examinaram a expressão acima de maneira tal que $\mathbf{y}$ representa um vetor resposta de um modelo linear no qual $(\mathbf{A}_1, \dots, \mathbf{A}_{k+1})$ e $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ correspondiam, respectivamente, a uma partição da matriz de planejamento e uma partição do vetor dos coeficientes de regressão, sendo $\mathbf{x}_{k+1}$ um vetor de discrepâncias de $\mathbf{y}$, com comportamento linear.

Em seguida, Dempster et al. (1977) assinalaram que, em um contexto de um modelo linear normal, os elementos $\mathbf{x}_{k+1}$ são comumente considerados como independentes e com distribuição $N(0, \sigma^2)$, conforme anteriormente assumido por meio de $\mathbf{x}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I})$.

Em contrapartida, em um contexto de (4.25) e uma vez que todos os elementos da partição do vetor dos coeficientes de regressão $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ admitem distribuição normal, Dempster et al. (1977) acabaram por converter um modelo de efeitos fixos em um de efeitos aleatórios.

Dessa maneira, as estatísticas suficientes (em um contexto de família exponencial) seriam

$$\mathbf{t}(\mathbf{x}) = (\mathbf{x}_1^t \mathbf{x}_1, \mathbf{x}_2^t \mathbf{x}_2, \dots, \mathbf{x}_{k+1}^t \mathbf{x}_{k+1}), \quad (4.26)$$

de modo que o E-passo exigiria o cálculo da esperança condicional de $\mathbf{t}_i = \mathbf{x}_i^t \mathbf{x}_i$, dado $\mathbf{y}$, a qual seria dada por

$$\mathbf{t}_i^{(p)} = \boldsymbol{\mu}_i^{(p)} \boldsymbol{\mu}_i^{(p)t} + tr \boldsymbol{\Sigma}_i^{(p)} \quad (4.27)$$

e o M-passo, produzindo estimativas de máxima verossimilhança de σ_i^2 , dado $\mathbf{t}_i^{(p)}$, seria dado por

$$\sigma_i^{2(p+1)} = \mathbf{t}_i^{(p)} / r_i, \quad (4.28)$$

para $i = 1, \dots, k + 1$.

De um modo geral, esse método é capaz de fornecer estimativas para $\sigma_1^2, \sigma_2^2, \dots, \sigma_{k+1}^2$ considerando $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ como valores faltantes.

Por fim, Dempster et al. (1977) chamam atenção, ainda, para o fato de que as estimativas obtidas por meio desse método são as mesmas quando comparadas com aquelas descritas em Thompsom (1971), Harville (1977) e Corbeil and Searle (1976), isto é, as estimativas de máxima verossimilhança de Dempster et al. também são estimativas de máxima verossimilhança restrita (REML).

4.3 O modelo HGLM

A classe de modelos HLM descrita na Seção 4.2 está bem adaptada para situações nas quais a variável resposta no Nível 1 é contínua. Mas, de acordo com Raudenbush e Bryk (2002, p. 291), existem duas suposições acerca da classe de modelos HLM que, se não atendidas, podem inviabilizar a aplicação dessa classe de modelos a um conjunto de dados. São elas:

1. Linearidade. No caso, o valor esperado da variável resposta não seria função linear dos coeficientes de regressão;
2. Normalidade. A distribuição normal para os efeitos aleatórios poderia não ser uma suposição realista.

Raudenbush e Bryk (2002, p. 291) ressaltam, ainda, que, em certos casos, essas suposições poderiam não ser atendidas nem mesmo com uma transformação na variável resposta. A título de ilustração, os autores citam duas situações, ambas com variável resposta de contagem: i) quando a variável resposta é binária (ou com distribuição de Bernoulli) e função de ligação *logit* e ii) quando a variável resposta admite valores inteiros e não negativos, para a qual atribuiremos distribuição de Poisson e uma função de ligação *log*.

Assim, é possível notar que a classe de modelos HLM não é adequada para a estimação de modelos hierárquicos para um conjunto de dados quando as suposições supramencionadas (linearidade e normalidade) não são atendidas. Nesse contexto, surge a classe de modelos hierárquicos HGLM (*Hierarchical Generalized Linear Model*).

A classe de modelos HGLM constitui uma extensão natural para a classe de modelos HLM, na medida em que torna possível atribuir outras distribuições de probabilidades (notadamente as discretas), bem como propor funções de ligação não lineares. Nessa linha de raciocínio, a classe de modelos HLM constitui um caso particular da classe HGLM, cuja variável resposta admite distribuição normal e função de ligação *identidade*.

Apesar de Raudenbush e Bryk (2002, p. 293) descrever alguns casos da classe de modelos HGLM (Bernoulli, Poisson, multinomial ordinal e multinomial nominal), serão apresentados, a seguir, apenas os casos de modelos HGLM com respostas binária (com função de ligação *logit*) e de Poisson (função de ligação *log*) no intuito de ilustrar melhor o funcionamento dessa classe de modelos.

4.3.1 O Modelo HGLM para Respostas Binárias

Consideraremos, para a i -ésima UE, o número Y_{ij} de sucessos em m_{ij} ensaios, sendo p_{ij} a probabilidade de sucessos em cada um dos ensaios. Logo, assumiremos

$$Y_{ij}|p_{ij} \sim \text{binomial}(m_{ij}, p_{ij}), \quad (4.29)$$

que, pelas propriedades da distribuição binomial, produz média e variância dadas, respectivamente, por

$$\mathbb{E}(Y_{ij}|p_{ij}) = m_{ij}p_{ij} \quad (4.30)$$

e

$$\text{Var}(Y_{ij}|p_{ij}) = m_{ij}p_{ij}(1 - p_{ij}). \quad (4.31)$$

No caso particular em que $m_{ij} = 1$, a variável de interesse, Y_{ij} , também pode ser caracterizada, vale dizer, descrita pela distribuição de Bernoulli (de parâmetro p_{ij}), uma vez que passa a admitir apenas valores 0 ou 1.

A função de ligação *logit*, bastante comum para esse tipo de situação, é dada por

$$\eta_{ij} = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right), \quad (4.32)$$

com η_{ij} representando o logaritmo da chance de sucesso.

A partir das estimativas dos coeficientes de regressão ($\beta's$) e algumas manipulações algébricas, estimativas de probabilidade podem ser obtidas a partir de

$$p_{ij} = \frac{1}{1 + \exp(-\eta_{ij})}. \quad (4.33)$$

Em um modelo com dois níveis hierárquicos, teremos

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{s_q} \gamma_{qs} W_{sj} + u_{qj}, \quad (4.34)$$

com o intercepto (β_{0j}) e cada um dos coeficientes β_{qj} , $q = 1, \dots, Q$ podendo ser função de variáveis pertencentes a um nível hierárquico mais elevado, W_{sj} , $s = 1, \dots, s_q$, com efeitos aleatórios u_{qj} normalmente distribuídos de médias zero e uma matriz de variância-covariância T .

O modelo pode incorporar níveis hierárquicos mais elevados. Por exemplo, para elaborar a estrutura do terceiro nível hierárquico, seria necessário descrever aquelas variáveis de Nível 2 que estariam em função de certas variáveis de Nível 3.

4.3.2 O Modelo HGLM para Respostas de Poisson

Consideremos o número Y_{ij} de ocorrências de um determinado evento e em um dado intervalo de tempo, m_{ij} , também conhecido como *exposição*. Raudenbush e Bryk (2002, p. 309) ilustra Y_{ij} como o número de crimes que um indivíduo i comete na vizinhança j durante o período de 5 anos, de modo que $m_{ij} = 5$.

Ao atribuir a distribuição de Poisson, isto é, ao supor que $Y_{ij}|\lambda_{ij}$ tenha distribuição de Poisson de parâmetro λ_{ij} levando, ainda, em consideração a exposição m_{ij} , temos, então, que a média e a variância serão dadas por

$$\mathbb{E}(Y_{ij}|\lambda_{ij}) = m_{ij}\lambda_{ij} \quad (4.35)$$

e

$$Var(Y_{ij}|\lambda_{ij}) = m_{ij}\lambda_{ij}, \quad (4.36)$$

uma vez que, pelas propriedades da distribuição de Poisson, a média é igual à variância. Em caso de $m_{ij} = 1$ (1 dia, 1 mês, 1 ano etc.), teremos a forma usual da distribuição de Poisson, com média e variância iguais a λ_{ij} .

A função de ligação em um contexto de regressão de Poisson é o logaritmo, de modo que

$$\eta_{ij} = \log(\lambda_{ij}). \quad (4.37)$$

A partir das estimativas dos coeficientes do modelo de regressão também é possível obter estimativas do logaritmo das taxas médias, $\hat{\eta}_{ij}$, sendo que, por meio de algumas manipulações algébricas, as taxas médias estimadas podem ser obtidas pela inversa da própria função de ligação, isto é, por meio de $\hat{\lambda}_{ij} = \exp(\hat{\eta}_{ij})$.

Ao considerar também um modelo hierárquico de 2 níveis, teremos a mesma estrutura dada em (4.34), com as mesmas suposições para os efeitos aleatórios e as mesmas possibilidades de extensão para modelos hierárquicos de níveis mais elevados.

4.3.3 Métodos de Estimação para HGLM

Conforme relatado na Seção 4.3, a classe de modelos HLM constitui caso particular da classe mais geral HGLM, com função de ligação *identidade* no nível N-1 e distribuição normal multivariada para os efeitos aleatórios nos níveis superiores.

De outro modo, a classe de modelos HGLM envolve:

- funções de ligação não lineares;
- variáveis resposta não restritas à distribuição normal;
- efeitos aleatórios com distribuição normal multivariada.

Logo, é possível atribuir, tanto para HLM quanto para HGLM, alguma função de verossimilhança para o processo de estimação de modelos, uma vez que a variável resposta pode assumir distribuição normal (HLM) ou outras distribuições de probabilidade (HGLM). Porém, de um ponto de vista mais técnico, problemas e desafios computacionais para a estimação de modelos HLM são, na prática, menos desafiadores.

A Tabela 4.1, a seguir, resume alguns dos métodos existentes de estimação de modelos HLM e HGLM, originalmente organizada por Raudenbush e Bryk (2002, p. 437).

Tabela 4.1 – Modelos, estimadores e algoritmos computacionais

<i>Models</i>	<i>Estimation Theory</i>	<i>Computational Algorithm</i>
Hierarchical linear models (two level)	Full ML Restricted ML Bayes	EM algorithm or Fisher scoring EM algorithm or Fisher scoring Gibbs sampling
Hierarchical linear models (three level) Cross-classified, multivariate models	Full ML	EM algorithm or Fisher scoring
Hierarchical generalized linear models (two- and three-level models)	Penalized quasi-likelihood Full ML	EM or Fisher scoring with iteratively reestimated weights Laplace approximation with Fisher scoring

Fonte: Raudenbush e Bryk (2002, p. 437)

Conforme é possível notar, para cada método de estimação, existem algoritmos computacionais específicos, dado o grau de sofisticação dessas classes de modelos.

É possível notar, ainda, na Tabela 4.1, o resultado alcançado na Seção 4.2.2, acerca do artigo de Dempster et al. (1977), a saber, de que as estimativas obtidas por esses autores por meio do algoritmo E-M, além de serem de máxima verossimilhança, também são de máxima verossimilhança restrita.

Um outro ponto que pode ser assinalado para a classe de modelos HGLM diz respeito ao método de estimação PQL, também utilizado para a estimação de modelos GLMM, já analisado no Capítulo 3.

Conforme constatado com a classe de modelos GLMM, o problema de cálculo de integrais também se verifica com a classe de modelos hierárquicos. Apenas para ilustrar esse fato, Raudenbush e Bryk (2002, p. 438) expressam $\mathbf{Y}$ como o vetor de todas as respostas em Nível 1, $\mathbf{u}$ como o vetor de todos os efeitos aleatórios contidos em Nível 2 (ou superior) e $\boldsymbol{\omega}$ como um vetor contendo todos os demais parâmetros a serem estimados (componentes de variância-covariância e coeficientes fixos de regressão). Logo, a função de verossimilhança dos dados, dados todos os parâmetros, é dada por

$$L(\mathbf{Y}|\boldsymbol{\omega}) = \int f(\mathbf{Y}|\mathbf{u}, \boldsymbol{\omega})p(\mathbf{u}|\boldsymbol{\omega})d\mathbf{u}, \quad (4.38)$$

com $f(\mathbf{Y}|\mathbf{u}, \boldsymbol{\omega})$ sendo a função densidade de probabilidade do vetor de respostas do Nível 1, dados os efeitos aleatórios e demais parâmetros e $p(\mathbf{u}|\boldsymbol{\omega})$ representando a distribuição dos efeitos aleatórios dados os demais parâmetros.

Em um contexto de algoritmo E-M, a maximização de (4.38) passaria a exigir duas etapas: i) a avaliação da integral e ii) a maximização da integral. De acordo com Raudenbush e Bryk (2002, p. 439), apesar de, em um contexto HLM, a integral (4.38) ser de fácil avaliação, a sua maximização ainda é complicada. Já para a classe HGLM, tanto o processo de integração quanto a maximização de (4.38) constituem tarefas complicadas.

Raudenbush e Bryk (2002, p. 455) explicam esse fato por meio do seguinte raciocínio: se, em (4.38), a função p fosse, em linguagem bayesiana, exatamente *a priori* conjugada de f (p. ex., para f binomial, p deveria representar uma distribuição beta ou, de modo diferente, para f Poisson, p deveria representar uma distribuição gama etc.), o problema seria apenas de maximização (e não mais de avaliação). Mas, dentre as distribuições de probabilidade envolvidas na integral (f e p), nem sempre p será *a priori* conjugada de f e, em razão disso, métodos numéricos são demandados para encontrar soluções.

De um modo geral, são vários os procedimentos de estimação para a classe de modelos HGLM, e Raudenbush e Bryk (2002) destacam vários artigos cujos autores têm aprimorado cada um desses procedimentos. Dentre esses procedimentos, os autores destacaram:

- Método PQL (p. 456);
- Método de simulação de Monte Carlo (p. 459);
- Método de aproximação de Laplace (p. 460).

4.4 A relação entre HGLM e GLMM

A relação entre as classes de modelos HGLM e GLMM está registrada na literatura. Conforme pode ser notado, foi destacado, no início da Seção 4.2, que a própria classe de modelos hierárquicos, na Biometria, é designada como *Modelos de Efeitos Mistos* e ou como *Modelos de Coeficientes Aleatórios*.

Breslow (2004), um dos responsáveis pelo desenvolvimento da classe de modelos GLMM, reconhece essa relação ao afirmar que:

The GLMM is a model for the hierarchical regression analysis of a series of n univariate response measurements y_i on a p -dimensional covariates x_j associated with fixed effects and q -dimensional covariates z_i associated with random effects of interest ($i = 1, \dots, n$). (p. 3)

A fim de explicitar essa relação, consideremos que um modelo HGLM geral no Nível 1 pode ser descrito por

$$Y_j = X_j\beta_j + r_j, \quad r_j \sim N(\mathbf{0}, \sigma^2 I), \quad (4.39)$$

sendo:

- Y_j um vetor $n_j \times 1$;
- X_j uma matriz $n_j \times (Q + 1)$ de variáveis preditoras;
- β_j um vetor $(Q + 1) \times 1$ de parâmetros desconhecidos;
- I a matriz identidade de ordem $n_j \times n_j$;
- r_j um vetor de erros aleatórios com média representada por um vetor $\mathbf{0}$ e matriz de variância-covariância com σ^2 na diagonal principal e todos os demais elementos iguais a 0.

No nível hierárquico imediatamente superior (Nível 2), o modelo HLM poderia assumir a forma geral

$$\beta_j = W_j \gamma + u_j, \quad u_j \sim N(\mathbf{0}, T), \quad (4.40)$$

sendo:

- W_j uma matriz $(Q + 1) \times F$ de variáveis preditoras;
- γ é um vetor $F \times 1$ de efeitos fixos;
- u_j é um vetor de erros ou efeitos aleatórios;
- T é uma matriz de variância-covariância relativa ao efeito aleatório u_j .

Ao substituir (4.40) em (4.39), teríamos o modelo agregado

$$\begin{aligned} Y_j &= X_j(W_j \gamma + u_j) + r_j \\ Y_j &= X_j W_j \gamma + X_j u_j + r_j. \end{aligned} \quad (4.41)$$

Nesse modelo resultante, todas as variáveis do Nível 1, isto é, aquelas contidas em X_j , estão associadas tanto com efeitos fixos quanto com efeitos aleatórios, por meio de γ e u_j , respectivamente.

Contudo, é possível reescrever esse modelo com base em uma premissa diferente, a saber, de que apenas algumas variáveis do Nível 1 (um subconjunto de X_j , especificamente designado por Z_j) estejam associadas com efeitos aleatórios. Em outros termos, teríamos o modelo

$$Y_j = X_j W_j \gamma + Z_j u_j + r_j, \quad (4.42)$$

com visível correspondência com a classe de modelos GLMM, analisada no Capítulo 3.

No prefácio do editor para a segunda edição de Raudenbush e Bryk (2002) também consta uma passagem que deixa mais clara essa relação: “*Perhaps even more importantly, mixed models (of which multilevel models are an important special case) have become very dominant in statistics.*” (p. xxiii)

Bates (2010) também afirma que:

Because each level of sample occurs with one and only one level of batch we say that sample is nested within batch. Some presentations of mixed-effects models, especially those related to multilevel modelling [Rasbash et al., 2000] or hierarchical linear models [Raudenbush e Bryk, 2002], leave the impression that one can only define random effects with respect to factors that are nested. This is the origin of the terms 'multilevel', referring to multiple, nested levels of variability, and 'hierarchical', also invoking the concept of hierarchy of levels. To be fair, those references do describe the use of models with random effects associated with non-nested factors, but such models tend to be treated as a special case. (p. 39)

Logo no parágrafo seguinte, Bates (2010) examina uma possível causa dessa aparente confusão:

The blurring of Mixed-effects models with the concept of multiple, hierarchical levels of variation results in an unwarranted emphasis on 'levels' when defining a model and leads to considerable confusion. It is perfectly legitimate to define models having random effects associated with non-nested factors. The reasons for the emphasis on defining random effects with respect to nested factors only are that such cases do occur frequently in practice and that some of the computational methods for estimating the parameters in the models can only be easily applied to nested factors. (p. 39)

5 O SOFTWARE HLM 7

5.1 Introdução

Neste capítulo, será introduzido um software estatístico especificamente desenvolvido para análise de dados hierárquicos, o HLM 7. Esse software estatístico específico será comparado com dois softwares estatísticos (R e SPSS) que, além de congregarem os métodos de estimação da classe de modelos hierárquicos, são de amplo alcance nas mais diversas comunidades científicas. Uma base de dados com aspectos hierárquicos e longitudinais será utilizada. À medida que os dados forem analisados (e, portanto, modelos são estimados), as estimativas dos parâmetros, bem como a forma de obtê-las, serão comparadas.

5.2 Características do software HLM 7

À medida que novas técnicas estatísticas foram sendo desenvolvidas ao longo do tempo, os softwares estatísticos também foram acompanhando esses desenvolvimentos e, em maior ou menor grau, passaram a incorporá-los em suas funcionalidades.

Atualmente existem alguns softwares estatísticos já bem estabelecidos e reconhecidos pela comunidade científica (SAS, SPSS, R etc.) que se destacam, também, por incorporar várias técnicas de estatística, no intuito de buscar atender aos profissionais das mais diversas áreas do conhecimento.

Entretanto, quando um profissional de uma determinada área que, em vista disso, trabalha mais diretamente com um conjunto mais específico de técnicas estatísticas, a escolha dentre softwares estatísticos “especializados” também constitui uma alternativa possível, principalmente quando dispõe de recursos e ou benefícios adicionais que podem ser de grande utilidade para esses profissionais.

A título de ilustração, softwares como SAS, SPSS e R abrangem uma grande variedade de técnicas estatísticas, as quais exigem um considerável esforço inicial desse profissional para que este passe a conhecer, entender e posteriormente utilizar seus comandos e ou funções básicas. Somente a partir dessa fase inicial é que o usuário poderá avançar, no sentido de entrar em contato, apreender e usar, com eficiência, os comandos e ou funções mais específicos e relacionados com sua área de pesquisa.

É interessante ressaltar que, no caso particular do software R, um número mínimo de comandos bem como alguma noção/conhecimento de programação são exigidos já desde o

início. Logo, uma certa quantidade de tempo e esforços adicionais para essa aprendizagem deverá ser despendida.

É precisamente nesse contexto que um software especializado pode, eventualmente, vir a ser uma alternativa mais interessante para profissionais e pesquisadores que lidam com um conjunto mais específico de técnicas estatísticas.

Para o caso específico de profissionais que lidam diretamente com problemas e situações que demandam uma abordagem com a classe de modelos hierárquicos, levando também em consideração o aspecto longitudinal das observações da variável resposta, o software HLM 7 constitui uma solução bastante interessante, principalmente por dispor de um recurso específico que pode ser considerado como um grande diferencial (quando comparado com os demais softwares).

De um modo geral, o HLM 7 (<http://www.ssicentral.com/hlm/>) pode ser considerado como um software “especializado” para estimar modelos de regressão hierárquicos, tendo sido desenvolvido (dentre outros) por S. W. Raudenbush e A. S. Bryk, autores qualificados nessa área de pesquisa.

Além disso, esse software dispõe de um recurso de grande utilidade: a tela HWLM. Essa tela proporciona um ambiente de trabalho extremamente profícuo, uma vez que permite ao usuário planejar, escolher e elaborar visual e manualmente modelos hierárquicos do seu interesse. Mais precisamente, na tela HWLM, o próprio usuário determina qual a variável resposta, quais variáveis independentes serão, de fato, utilizadas no nível 1 e nos demais níveis hierárquicos, se as variáveis devem ser centradas (na média) ou não, quais coeficientes (inclusive o intercepto) serão utilizados para modelar os efeitos hierárquicos em níveis mais elevados, se efeitos aleatórios serão levados, ou não, em consideração, além de poder estabelecer o próprio número de níveis hierárquicos do modelo (máximo de 4).

É importante notar que outros softwares também são capazes de estimar modelos hierárquicos, mas, conforme pode ser constatado, as saídas dos softwares, especificamente em modelos hierárquicos, podem variar de software para software. Apenas a título de ilustração, a biblioteca `lme4` do software R não oferece em suas saídas as estimativas do *p-valor*, sejam para as estimativas dos efeitos fixos, sejam para as estimativas dos efeitos aleatórios. No entanto, a biblioteca `nlme` (também do software R) fornece as quantidades de *p-valor* somente para as estimativas dos efeitos fixos.

Outros softwares podem, também, apresentar diferenças não somente em suas saídas mas nas próprias estimativas dos parâmetros do modelo. Isso ocorre com certa frequência em modelos computacionalmente mais sofisticados, e um dos motivos para esse tipo de ocorrência

reside no fato de haver diferenças entre os algoritmos computacionais utilizados por diferentes empresas de softwares.

Contudo, ao chamar a atenção para essas questões, desejamos, na realidade, ressaltar que não existe garantia de uniformidade entre os softwares estatísticos em termos das informações fornecidas (saídas), sejam elas quantitativas (estimativas) ou qualitativas (medidas estatísticas envolvidas).

Com relação a esse recurso apontado do HLM 7, isto é, a tela HWLM, são vários os aspectos interessantes que poderiam ser ressaltados, mas, especificamente, dois deles merecem destaque especial:

- visualização clara e explícita da forma do modelo;
- facilidade de operacionalização.

A facilidade de operacionalização e ou da manipulação das informações com o próprio software para estimação de modelos hierárquicos constitui, por si só, um grande atrativo, uma vez que tempo e esforços necessários para dominá-lo são bastante reduzidos.

Com relação à primeira qualidade, ao permitir uma visualização contínua, clara e inteligível do modelo que o usuário deseja estimar, a tela HWLM propicia um ambiente (para o usuário) de total controle de todos os aspectos fundamentais do próprio modelo a ser estimado, tais como: i) escolha e determinação de cada uma das variáveis em seus respectivos níveis hierárquicos e ii) determinação do número de níveis hierárquicos no modelo, não disponível em nenhum outro software, além de dispensar a necessidade de conhecimentos de programação.

Da perspectiva dos outros softwares, e no tocante à classe de modelos hierárquicos, o fato de que algumas variáveis são modeladas em níveis hierárquicos mais elevados acaba por exigir uma atenção ainda mais pormenorizada por parte do usuário, enquanto que, da perspectiva do HLM 7, a tela HWLM, ao dispor desse recurso, torna o próprio processo de elaboração de um modelo hierárquico mais intuitivo.

A título de ilustração, vamos utilizar o conjunto de dados HS&B (High School & Beyond). Essa base de dados foi utilizada pelos autores do manual do software, isto é, por Raudenbush, Bryk, Cheong, Congdon e du Toit (2011, p. 15), tendo sido também, de acordo com eles, utilizada no Capítulo 4 de Raudenbush e Bryk (2002).

Raudenbush e Bryk (2002, p. 68) declaram que esses dados se referem a uma subamostra de uma pesquisa de 1982, a qual inclui 7.185 estudantes do ensino médio contidos em 160 escolas (90 públicas e 70 católicas). Raudenbush e Bryk (2002, p. 95) mencionam, ainda, que esses dados encontram-se disponíveis no Departamento de Educação dos EUA (*Office of Educational Research and Improvement, U.S. Department of Education*).

Vamos supor que desejamos estimar um modelo hierárquico de dois níveis para as notas de matemática de alunos de uma determinada série escolar. No caso, acredita-se que essas notas sejam descritas por um intercepto e que sejam função do nível socioeconômico de cada aluno (SES), caracterizando, desse modo, a equação de nível 1.

No nível 2, apenas por suposição, acredita-se que a nota média de cada escola seja influenciada pelo tipo de escola (SECTOR) e que o nível socioeconômico de cada aluno, SES, seja explicado pelo nível socioeconômico médio dos alunos da escola (MEANSES).

Com esse modelo em mente, a sua elaboração, bem como sua efetiva estimação, podem ser realizadas com o apoio da tela HWLM, à medida que introduzimos passo a passo essas informações, com o acompanhamento do próprio processo de elaboração das equações do modelo.

Ao acessar o ambiente da tela HWLM para esse conjunto de dados, o software HLM 7 fornecerá a tela dada pela Figura 5.1.

Figura 5.1 – Tela HWLM

Fonte: O autor.

No primeiro nível hierárquico, de acordo com o modelo pretendido, a variável resposta MATHACH é função do nível socioeconômico (SES). O estabelecimento dessa primeira relação pode ser observada na Figura 5.2. Note que, ao determinar o primeiro nível hierárquico, o próprio software permite visualizar as possibilidades para o segundo nível hierárquico.

Figura 5.2 – Tela HWLM: Determinação do primeiro nível hierárquico

Fonte: O autor.

Já no segundo nível hierárquico, uma vez que, nesse modelo exemplificado, o nível das notas de Matemática (MATHACH) é influenciado pelo tipo de escola (SECTOR), enquanto que o nível socioeconômico pode ser explicado pelo nível socioeconômico médio dos alunos da escola (MEANSES), a Figura 5.3 descreve esse processo de “modelagem” na tela HWLM.

Figura 5.3 – Tela HWLM: Exemplo de modelo hierárquico de dois níveis

Fonte: O autor.

Note que, uma vez que o modelo é, por suposição, de dois níveis, a própria Figura 5.3 constitui o modelo exemplificado. É importante ressaltar, também, que os termos aleatórios no nível mais elevado, u_{0j} e u_{1j} , estão “ativados” na tela HWLM, de modo que esse modelo é tanto de intercepto quanto que de coeficiente aleatórios. Mas isso pode ser alterado pelo pesquisador. Se o interesse estivesse baseado em, por exemplo, um modelo de coeficiente aleatório, então, um simples “clique” no termo aleatório u_{0j} tornaria indisponível esse termo, produzindo, por conseguinte, um modelo de coeficiente aleatório. O mesmo raciocínio é válido para um modelo de coeficiente aleatório.

De uma perspectiva mais ampla, o HLM 7 é capaz de lidar com estruturas hierárquicas mais complexas, tendo suas funcionalidades distribuídas em 18 capítulos.

No Capítulo 1, os principais conceitos estatísticos necessários para a estimação de modelos hierárquicos de dois níveis são apresentados: estimadores *Empirical Bayes* (EB), para os coeficientes de nível 1, estimadores de Mínimos Quadrados Generalizados (GLS), para estimadores de nível 2, estimadores de Máxima-Verossimilhança para componentes de variância-covariância, testes de hipóteses, estimação de Máxima-Verossimilhança em contraposição ao método de estimação de Máxima-Verossimilhança Restrita e GEE, tendo, este último, sido incorporado ao software apenas na condição de método de estimação.

Do segundo ao sexto capítulos, os modelos hierárquicos de níveis 3 (HLM3) e 4 (HLM4) são abordados, cada qual contando com um capítulo próprio de exemplos.

No Capítulo 7, os autores introduzem a classe de modelos *Hierarchical Generalized Linear Models* (HGLM), com possibilidade de até quatro níveis. O que o diferencia dos anteriores, é que HGLM permite que as variáveis respostas sejam binárias, de contagem e multinomiais (nominais ou ordinais). Na prática, implica uma extensão similar àquela produzida pela classe de modelos GLM com relação aos modelos que admitiam resposta com uma única distribuição (normal).

No Capítulo 9, os autores introduzem a classe de modelos hierárquicos multivariados (HMLM) que, segundo eles, são modelos que permitem realizar análises de medidas repetidas multivariadas, no sentido de que as observações repetidas estão aninhadas em uma mesma UE (Raudenbush et al., 2011, p. 147).

No Capítulo 10, os autores introduzem a classe de modelos HMLM2 que, basicamente, permite levar em consideração múltiplas respostas para unidades que se encontram aninhadas em unidades hierarquicamente superiores e realizam exemplos de aplicação.

No Capítulo 11, alguns problemas específicos tais como variáveis latentes e dados faltantes são abordados para a correta aplicação no software.

O Capítulo 12 aborda a classe *Hierarchical Cross-classified Random Effects Model* (HCM2), tendo como objetivo principal lidar com as situações nas quais os dados não se encontram perfeitamente aninhados.

No Capítulo 13, são apresentados exemplos ilustrativos para os modelos HCM2, sendo que, nos Capítulos 14 e 15, a classe de modelos HCM3 é apresentada com alguns exemplos ilustrativos.

No Capítulo 16, a classe de modelos *Hierarchical Linear Model with Cross-classified Random Effects* (HLMHCM) é apresentada e, a título de ilustração, é dado o exemplo de um modelo de crescimento para um grupo de crianças que moram em um mesmo bairro e frequentam uma mesma escola, de forma que as unidades do nível 1 (medidas repetidas)

encontram-se aninhadas no nível 2 (crianças), com unidades de nível 2 cruzadas por linhas (bairros) e colunas (escolas).

No Capítulo 17, são apresentados exemplos práticos para HLMHCM, e o Capítulo 18 ficou reservado para os aspectos gráficos.

A seguir, será analisado um banco de dados com características hierárquicas e longitudinais por meio do software HLM 7.

5.3 Análise de um banco de dados

Existem várias bases de dados que acompanham o software HLM 7, tendo, dentre elas, algumas de natureza hierárquica e longitudinal, simultaneamente. Em geral, alguns desses conjuntos de dados encontram-se divididos em arquivos distintos, cada qual com variáveis pertencentes a um determinado nível hierárquico. Por exemplo, EG1.SAV, EG2.SAV e EG3.SAV dizem respeito a um conjunto de dados com 7.242 observações obtidas de 1.721 crianças acompanhadas da primeira até a sexta série com variáveis de três níveis hierárquicos. Já os arquivos NYS1.SAV e NYS2.SAV dizem respeito a um conjunto de dados com 1.079 observações obtidas de entrevistas com 239 jovens de 11 anos de idade por 5 anos consecutivos (com início em 1976). Em NYS1.SAV, encontram-se as variáveis do nível 1 e, em NYS2.SAV, as variáveis de nível 2.

No manual do software HLM 7, esses arquivos são utilizados em alguns capítulos no intuito de ilustrar a técnica (ou conjunto de técnicas) em questão.

Contudo, seria interessante avaliar o desempenho do software com um outro banco de dados, isto é, com um banco de dados que não faz parte dos recursos do HLM 7. Dessa forma, optou-se por utilizar uma base de dados também com características hierárquicas e longitudinais, do SPSS, abordada em Heck, Thomas e Tabata (2014).

Esse banco de dados diz respeito a um estudo sobre a proporção dos graduados em instituições públicas americanas de ensino superior ao longo de 11 anos (aspecto longitudinal), levando em consideração os estados (aspecto hierárquico) a que pertencem as instituições analisadas.

No nível 1, encontram-se medidas repetidas da proporção de estudantes de graduação que se formaram em uma determinada instituição ao longo do tempo. No nível 2, algumas variáveis relativas às características das instituições estão presentes no intuito de se detectar possíveis diferenças nas taxas de formandos entre as instituições. No nível 3, encontram-se duas variáveis (padronizadas) no nível dos estados: i) *aveRetention* e ii) *aveFamilyshare*. A primeira mede o

percentual de calouros (no estado), enquanto que a segunda variável mede um percentual que reflete a contribuição das famílias para com a educação superior no estado.

Seguiremos os mesmos procedimentos adotados por Heck et al. (2014), no que se refere ao processo de estimação de modelos hierárquicos, a saber, de modelos mais simples para aqueles com maior grau de complexidade. Todos esses procedimentos encontram-se descritos no Capítulo 6 da obra dos autores.

No início do capítulo, os autores sugerem três perguntas de pesquisa:

- 1) Qual o nível médio e a forma das trajetórias das proporções de graduação das instituições ao longo do tempo?
- 2) Quais variáveis intrainstitucionais e interinstitucionais explicam os níveis das proporções de graduação?
- 3) Como as trajetórias das proporções de graduação variam de acordo com as características de cada estado?

De forma a responder esses questionamentos, os autores recorrem à seguinte estratégia de três níveis:

- No nível 1, possíveis relações envolvendo variáveis variantes no tempo (*time-varying*) que podem explicar as proporções das graduações serão analisadas.
- No nível 2, informações no nível institucional serão analisadas de forma a se avaliar e medir a influência que podem vir a exercer sobre as taxas de graduação ao longo do tempo.
- No nível 3, os efeitos de diferentes padrões de comportamentos econômicos e políticos regionais (nesse caso, estaduais) sobre as taxas de graduação das instituições observadas poderão ser avaliados.

Os autores denominam essa estratégia como *three-level growth formulation*.

Na sequência, serão elaborados cinco modelos (sempre partindo do mais simples para o mais sofisticado) com o apoio de três softwares: HLM 7, SPSS e R. Isso permitirá comparar as diferentes formas de uso e funcionamento desses softwares, bem como controlar os resultados já conhecidos pelo SPSS em Heck et al. (2014).

Com relação ao software R, foi escolhida a biblioteca `nlme`. Uma outra possibilidade seria a biblioteca `lme4`, a qual permite estimar modelos generalizados, isto é, permitindo outros modelos de distribuição de probabilidades para a variável resposta. Contudo, não possui a mesma versatilidade no que se refere à estrutura de correlação. Assim, pelo fato de Heck et al.

(2014) terem assumido uma estrutura de correlação autorregressiva de ordem 1 (AR1), a biblioteca de referência para estimar modelos hierárquicos no R será nlme.

Para a estimação dos modelos a seguir, o método de máxima verossimilhança restrita (REML) será utilizado em todos os casos. Todas as saídas do software HLM 7 encontram-se no apêndice desse capítulo.

5.3.1 Modelo 1: Modelo Nulo

Por uma questão de simplificação de nomenclaturas, utilizaremos as utilizadas pelo software HLM 7, até mesmo para facilitar a identificação dos parâmetros dos modelos. Além disso, ao informar, durante o processo de imputação dos dados, que o modelo desejado é um modelo linear hierárquico de nível 3 com medidas repetidas no nível 1, as nomenclaturas dos parâmetros (bem como os subíndices de cada um dos parâmetros e ou coeficientes) sofrerão algumas modificações para levar em consideração essas características.

Um modelo nulo (ou incondicional) é um modelo sem variáveis preditoras. Esse tipo de modelo permite particionar a variância da média global em duas partes:

- 1) Componentes intraindividuais;
- 2) Componentes entre os indivíduos.

No caso, ambos os componentes serão utilizados para calcular o ICC (*Intraclass Correlation*). Para o nível 1, o ICC do nível 1 é dado por

$$\rho_1 = \frac{\sigma_{n1}^2}{\sigma_{n1}^2 + \sigma_{n2}^2 + \sigma_{n3}^2}, \quad (5.1)$$

sendo $\sigma_{n1}^2, \sigma_{n2}^2$ e σ_{n3}^2 os componentes de variância (estimados) dos níveis 1, 2 e 3, respectivamente. Assim, ρ_1 pode ser interpretado como a proporção de variabilidade contida nos resultados do nível 1. A mesma interpretação pode ser estendida para os demais níveis.

Para os níveis 2 e 3, os ICC desses níveis serão dados por

$$\rho_2 = \frac{\sigma_{n2}^2}{\sigma_{n1}^2 + \sigma_{n2}^2 + \sigma_{n3}^2} \quad (5.2)$$

e

$$\rho_3 = \frac{\sigma_{n3}^2}{\sigma_{n1}^2 + \sigma_{n2}^2 + \sigma_{n3}^2}, \quad (5.3)$$

respectivamente.

Com relação à estrutura do modelo, y_{mti} representa a variável resposta observada na ocasião t , para a i -ésima instituição. Logo, a expressão para o modelo nulo, especificamente no nível 1, será dada por

$$y_{mti} = \psi_{0ti} + \varepsilon_{mti}. \quad (5.4)$$

No nível 2, o intercepto assumirá a forma

$$\psi_{0ti} = \pi_{00i} + e_{0ti}. \quad (5.5)$$

No nível 3, teremos

$$\pi_{00i} = \beta_{000} + r_{00i}. \quad (5.6)$$

Os resultados estimados para o modelo nulo com os três softwares (o resultado do SPSS foi apenas transcrito) encontram-se na Tabela 5.1.

Por meio da Tabela 5.1, é possível notar que as estimativas obtidas pelos três softwares ficaram bem próximas. Com base nessas estimativas da Tabela 5.1, os ICC puderam ser avaliados e encontram-se dispostos na Tabela 5.2.

Tabela 5.1 – Saídas das estimativas para o modelo Nulo

		<i>Softwares</i>		
		HLM 7	R	SPSS
<i>Efeitos fixos</i>	Intercepto (β_{000})	0,431584	0,4311116	0,431
	Erro – nível 1 (ε_{mti})	0,00340	0,004105	0,004
<i>Efeitos Aleatórios</i>	Erro – nível 2 (e_{0ti})	0,01703	0,0163969	0,016
	Erro – nível 3 (r_{00i})	0,00319	0,003254	0,003

Fonte: O autor.

Uma vez que as estimativas dos efeitos aleatórios da Tabela 5.1 também ficaram próximas, os valores dos ICCs, que dependem daquelas, também tendem, de certa forma, a ficarem próximos.

Tomando como referência o software HLM 7, as estimativas da Tabela 5.2 sugerem que a proporção da variância em y_{mti} (taxas de graduação) entre os estados (nível 3) é de cerca de 13,51%. Com relação ao ICC do nível 2, seu valor sugere que 72,10% da variância em y_{mti} são explicados pelas diferenças entre as instituições e, de modo semelhante, 14,39% são o percentual da variância no nível das instituições.

Tabela 5.2 – Estimativas dos ICCs para o modelo Nulo

	<i>Softwares</i>		
	HLM 7	R	SPSS
ICC – nível 1	0,1439	0,1727	0,173
ICC – nível 2	0,7210	0,6902	0,690
ICC – nível 3	0,1351	0,1369	0,137

Fonte: O autor.

Esses valores apontam para uma variabilidade considerável nos três níveis hierárquicos, o que torna a análise multinível adequada.

A Figura 4 mostra a tela HWLM elaborada para o modelo nulo, com a opção “*Mixed*” acionada (canto inferior direito) para sua exibição, descrito algebricamente por uma única equação.

A obtenção dessa expressão de uma única equação é obtida por sucessivas substituições, isto é, pela substituição da(s) equação(ões) do(s) nível(is) mais elevado(s) na(a) equação(ões) do(s) nível(is) imediatamente menos elevado(s).

No caso do modelo nulo, ao substituírmos a equação do nível 3 na equação do nível 2 e esse resultado na equação do nível 1, teremos

$$y_{mti} = \beta_{000} + r_{00i} + e_{0ti} + \varepsilon_{mti}, \quad (5.7)$$

coincidindo com a expressão descrita na parte inferior da tela HWLM.

Figura 5.4 – Tela HWLM da saída do modelo Nulo

Fonte: O autor.

Para a obtenção das saídas do software R (pela biblioteca *nlme*), o comando dado pela Tabela 5.3 deverá ser executado.

Tabela 5.3 – Comando executado pelo R para o modelo Nulo

```
fit_null <- lme(gradproportion ~ 1, random = ~ 1 | stateid/rid, correlation = corAR1(form = ~ 1 | stateid/rid))
```

Fonte: O autor.

De acordo com o comando da Tabela 5.3, *gradproportion* é a variável resposta de um modelo de intercepto aleatório com componentes aleatórios nos níveis 2 e 3 e estrutura de correlação autorregressiva de ordem 1 (AR1).

Para a devida exibição (saída) do modelo, o comando `summary(fit_null)` deverá ser executado, sendo que as estimativas dos componentes de variância estarão sob a forma de desvio padrão, ou seja, será necessário, ainda, elevar esses valores ao quadrado.

No caso do SPSS, o usuário terá que acessar algumas telas tendo que, em cada uma delas, fornecer algumas informações. A título de ilustração, apenas os comandos executados pelo SPSS, fornecidos no log, serão apresentados. Estes encontram-se disponíveis na Tabela 5.4.

Tabela 5.4 – Comandos executados pelo SPSS para o modelo Nulo

```
GET
  FILE='C:\Users\Koki\Econometria\Dados_Longitudinais\Livro
IBM\ch6multivarML.sav'.
DATASET NAME ConjuntodeDados1 WINDOW=FRONT.
MIXED gradproportion
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=| SSTYPE(3)
  /METHOD=REML
  /PRINT=G SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(stateid) COVTYPE(ID)
  /RANDOM=INTERCEPT | SUBJECT(rid*stateid) COVTYPE(ID)
  /REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).
```

Fonte: O autor.

Uma vez que os resultados sugerem uma abordagem multinível, um modelo com um grau maior de sofisticação será estimado a seguir.

5.3.2 Modelo 2: Adicionando Taxas de Crescimento

Heck et al. (2014, p. 252) chamam a atenção para o fato de que, quando os períodos de tempo são relativamente curtos e quando não existem muitas observações no nível individual (no caso, das instituições), o modelo de tendência linear tende a ser uma boa aproximação. Mas como os autores ainda não estavam totalmente convencidos dessa possibilidade, introduziram, também, uma tendência quadrática e conduziram uma rápida avaliação comparando ambas as tendências (linear e quadrática), somente por meio dos efeitos fixos.

Heck et al. (2014, p. 254) decidiram utilizar uma codificação diferente para as tendências linear e quadrática, a saber, *time1* e *quadtime1*. Isso porque os autores desejavam uma interpretação específica para os efeitos fixos e/ou aleatórios ao longo do tempo (Heck et al., 2014, p. 254). No caso de *time1*, em vez de considerar a própria sequência de 11 anos do tempo, *time* = 1, 2, ..., 11, a codificação *time1* = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, também com 11 períodos para mensuração, será utilizada. Segundo os autores, a vantagem dessa codificação consiste em viabilizar análises para as mudanças no crescimento das proporções de graduação dentro do intervalo [0,1], para todo o período do estudo. O mesmo se aplica para a tendência quadrática. Em vez de considerar *quadtime*, isto é, o quadrado dos 11 valores de *time* (0, 1, 4, 9, 16, 25, 36, 49, 64, 81, 100), a codificação *quadtime1*, com valores 0, 0.01, 0.04, 0.09, 0.16, 0.25, 0.36, 0.49, 0.64, 0.81, 1, será utilizada, também no intuito de analisar as mudanças dentro do intervalo [0,1]. Em ambos os casos (*time1* e *quadtime1*), o valor inicial é zero. Com isso, o intercepto passa a ser interpretado como o *status* inicial da proporção de graduação de determinada instituição (nível 2) pertencente a determinado estado (nível 3).

Na Tabela 5, encontram-se os resultados das estimativas de *time1* e *quadtime1*, de modo a realizar uma comparação.

Tabela 5.5 – Análise comparativa entre as tendências linear e quadrática (continua)

	<i>Software</i>					
	HLM 7		R		SPSS	
	est.	p-valor	est.	p-valor	est.	p-valor
Intercepto	0,4123	< 0,001	0,4150	<2e-16	0,429	0,000
<i>time1</i>	0,0324	0,030	0,03313	0,125	0,039	0,002
<i>quadtime1</i>	0,0083	0,542	0,007059	0,734	0,013	0,197

Fonte: O autor

As estimativas, mais uma vez ficaram todas bem próximas. Os softwares HLM 7 e SPSS apresentaram um mesmo padrão, a saber, com os coeficientes de intercepto e *time1* estatisticamente significantes, sugerindo que a tendência linear ajusta-se melhor ao conjunto de dados.

É importante destacar que, pelo software R, nenhuma das tendências mostrou-se estatisticamente significativa ao nível de 0,05. Mas, ao levar em consideração os resultados do SPSS e do HLM 7, será introduzida apenas a tendência linear nessa etapa.

Dessa maneira, a forma do modelo 2 passa a ser expressa por

$$y_{mti} = \psi_{0ti} + \psi_{1ti}time1 + \varepsilon_{mti}. \quad (5.8)$$

No nível 2, o intercepto e o coeficiente serão dados, respectivamente, por

$$\psi_{0ti} = \pi_{00i} + e_{0ti} \quad (5.9)$$

e

$$\psi_{1ti} = \pi_{10i} + e_{1ti}. \quad (5.10)$$

No nível 3, teremos

$$\pi_{00i} = \beta_{000} + r_{00i} \quad (5.11)$$

e

$$\pi_{10i} = \beta_{100} + r_{10i}. \quad (5.12)$$

Uma importante observação deve ser feita. Apesar de Heck et al. (2014, p. 253-254) explicitamente atribuírem (nas expressões do modelo) termos aleatórios apenas para os interceptos (níveis 2 e 3), as saídas do modelo acabaram fornecendo estimativas dos termos aleatórios para o coeficiente da tendência linear em ambos os níveis (2 e 3).

Dessa maneira, vamos considerar, para efeito de análise comparativa, um modelo com termos aleatórios tanto no intercepto quanto nos coeficientes, em ambos os níveis hierárquicos superiores (2 e 3), uma vez que, na prática, os autores acabaram estimando essas quantidades.

Além disso, esse mesmo procedimento será replicado em todos os modelos a serem, ainda, estimados neste capítulo, isto é, atribuir termos aleatórios em ψ_{0ti} , ψ_{1ti} , π_{00i} e π_{10i} .

As estimativas do modelo 2 encontram-se na Tabela 5.6.

Para a obtenção das saídas do software R, o comando dado pela Tabela 5.7 deve ser executado.

Tabela 5.6 – Saídas das estimativas para o modelo 2

Efeito		Software		
		HLM 7	R	SPSS
Fixo	Intercepto (β_{000})	0,411065	0,411032	0,411
	time1 (β_{100})	0,040737	0,0396326	0,040
	Erro – nível 1 (ε_{mti})	0,00234	0,002711	0,003
Aleatório	Erro – nível 2 (e_{0ti}) (intercepto)	0,01726	0,016605	0,016
	Erro – nível 2 (e_{1ti}) (time1)	0,00771	0,0057116	0,005
	Erro – nível 3 (r_{00i}) (intercepto)	0,00282	0,002902	0,003
	Erro – nível 3 (r_{10i}) (time1)	0,00035	0,0003408	0,000

Fonte: O autor.

De acordo com o comando da Tabela 5.7, *gradproportion* é a variável resposta de um modelo de intercepto e coeficiente aleatórios com componentes aleatórios nos níveis 2 e 3 e estrutura de correlação autorregressiva de ordem 1 (AR1).

Tabela 5.7 – Comando executado pelo R para o modelo 2

```
fit2 <- lme(gradproportion ~ time1, random = list(stateid = ~ 1 + time1, rid = ~ 1 + time1), correlation = corAR1(form = ~ 1 | stateid/rid))
```

Fonte: O autor.

Para a devida exibição (saída) do modelo no R, o comando `summary(fit2)` deverá, ainda, ser executado, sendo que as estimativas dos componentes de variância também estarão sob a forma de desvio padrão, ou seja, será necessário, ainda, elevar esses valores ao quadrado. Para o SPSS, os comandos executados e fornecidos no log encontram-se na Tabela 5.8.

Tabela 5.8 – Comandos executados pelo SPSS para o modelo 2

```
MIXED gradproportion WITH time1
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
SINGULAR(0.0000000000001) HCONVERGE(0,ABSOLUTE) LCONVERGE(0, ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
/FIXED=time1 | SSTYPE(3)
/METHOD=REML
/PRINT=G SOLUTION TESTCOV
/RANDOM=INTERCEPT time1 | SUBJECT(stateid) COVTYPE(DIAG)
/RANDOM=INTERCEPT time1 | SUBJECT(rid*stateid) COVTYPE(DIAG)
/REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).
```

Fonte: O autor.

Para a elaboração do modelo 2 pelo HLM 7, foi necessário apenas inserir TIME1 e executar o comando “Run Analysis” disponível na própria tela HWLM, visível na Figura 5. Além disso, a simples substituição das equações de nível 3 nas equações de nível 2 e 1 produzirá a expressão descrita na parte inferior da Figura 5 (*Mixed Model*).

Uma comparação entre as Tabelas 5.7, 5.8 e a Figura 5, é possível notar que o método de elaboração de modelos pelo HLM 7 dispensa a necessidade de programação (softwares R e eventualmente SPSS) e/ou de acesso e passagem por várias telas (característica central do SPSS). De modo diverso, a tela HWLM dispõe de um aspecto visual mais intuitivo, quando avaliado da perspectiva de um usuário.

Figura 5.5 – Tela HWLM da saída do modelo 2

Fonte: O autor.

Com relação à interpretação das estimativas do modelo 2, é possível constatar, de um modo geral (e por meio da Tabela 5.6), que existe uma forte proximidade entre as estimativas fornecidas pelos três softwares. Além disso, temos que:

- tanto o coeficiente da variável *time1* quanto o intercepto mostraram-se estatisticamente significantes ao nível de 0,05, em todos os três softwares;
- com relação ao intercepto (β_{000}), temos que, no *status* inicial ($t = 0$), a proporção de graduação por parte das instituições foi da ordem de 0,411 (41,1%);

- com relação ao coeficiente de *time1* (β_{100}), temos que as proporções de graduação (em nível estadual) cresceram no período considerado por volta de, aproximadamente, 4%, a partir do nível basal (intercepto) de 0,4111.

Com relação aos efeitos aleatórios no nível dos estados (nível 3), temos que:

- a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de que não existe, no nível dos estados, variação entre as proporções de graduação no *status* inicial do estudo (intercepto) foi rejeitada (ao nível de 0,05), sugerindo uma variação da ordem de 0,00282;
- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível dos estados também foi rejeitada, sugerindo uma taxa crescimento nas proporções de graduação entre os estados em torno de 0,00035.

Com relação aos efeitos aleatórios no nível das instituições (nível 2), temos que:

- a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de inexistência de variação no *status* inicial entre as proporções de graduação no nível das instituições foi rejeitada (ao nível de 0,05), sugerindo uma variação de, aproximadamente, 1,73% entre as instituições;
- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível das instituições também foi rejeitada, sugerindo que as proporções de graduação entre instituições cresceram em torno de 0,77%.

5.3.3 Modelo 3: Adicionando *time-varying covariates*

No modelo 3, serão introduzidas algumas covariáveis variantes no tempo. Logo, a forma assumida para o modelo 3 será dada por

$$y_{tij} = \psi_{0ti} + \psi_{1ti}time1_{mij} + \psi_{2ti}percentFinAid_{mti} + \psi_{3ti}tuition_{mti} + \varepsilon_{mti}, \quad (5.13)$$

sendo *percentFinAid* a porcentagem de estudantes que recebem auxílio financeiro e *tuition*, uma medida do nível médio dos custos com matrícula da instituição, ambas padronizadas e pertencentes ao nível hierárquico mais fundamental (nível 1).

Nos níveis 2 e 3, as equações permanecem inalteradas, isto é, idênticas àquelas já apresentadas em (5.9), (5.10), (5.11) e (5.12). Apenas para fins de exibição, seguem essas equações na mesma ordem.

$$\psi_{0ti} = \pi_{00i} + e_{0ti} \quad (5.9)$$

$$\psi_{1ti} = \pi_{10i} + e_{1ti} \quad (5.10)$$

$$\pi_{00i} = \beta_{000} + r_{00i} \quad (5.11)$$

$$\pi_{10i} = \beta_{100} + r_{10i} \quad (5.12)$$

As estimativas dos parâmetros do modelo 3 encontram-se na Tabela 5.9.

Tabela 5.9 – Saídas das estimativas para o modelo 3

Efeito		Software		
		HLM 7	R	SPSS
Fixo	Intercepto (β_{000})	0,475789	0,470922	0,469
	time1 (β_{100})	0,014657	0,0155076	0,016
	percentFinAid (ψ_{2ti})	0,001444	0,0015114	0,002
	tuition (ψ_{3ti})	0,054975	0,0511153	0,049
	Erro – nível 1 (ε_{mti})	0,00231	0,002647	0,003
Aleatório	Erro – nível 2 (e_{0ti}) (intercepto)	0,01657	0,016016	0,015
	Erro – nível 2 (e_{1ti}) (time1)	0,00775	0,0058637	0,005
	Erro – nível 3 (r_{00i}) (intercepto)	0,00249	0,0025941	0,003
	Erro – nível 3 (r_{10i}) (time1)	0,00030	0,0002909	0,000

Fonte: O autor.

Para a obtenção das saídas do software R, o comando dado pela Tabela 5.10 deverá ser executado.

Tabela 5.10 – Comando executado pelo R para o modelo 3

```
fit3 <- lme(gradproportion ~ time1 + percentFinAid + tuition, r
random = list(stateid = ~ 1 + time1, rid = ~ 1 + time1), correla
tion = corAR1(form = ~ 1 | stateid/rid))
```

Fonte: O autor.

De acordo com o comando da Tabela 5.10, *gradproportion* é a variável resposta de um modelo de intercepto e coeficiente aleatórios com componentes aleatórios nos níveis 2 e 3 e estrutura de correlação autorregressiva de ordem 1 (AR1). Para a devida exibição (saída) do modelo, o comando `summary(fit3)` deverá ser executado, com estimativas sob a forma de desvio padrão.

Os comandos do log do SPSS para o modelo 3 encontram-se na Tabela 5.11.

Para a elaboração do modelo 3 pelo HLM 7, foi necessário reproduzir a tela descrita pela Figura 6. A simples substituição das equações de nível 3 nas equações de nível 2 e 1 também produzirá a expressão descrita na parte inferior da Figura 6 (*Mixed Model*).

Tabela 5.11 – Comandos executados pelo SPSS para o modelo 3

```

MIXED gradproportion WITH time1 percentFinAid tuition
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=time1 percentFinAid tuition | SSTYPE(3)
  /METHOD=REML
  /PRINT=G SOLUTION TESTCOV
  /RANDOM=INTERCEPT time1 | SUBJECT(stateid) COVTYPE(DIAG)
  /RANDOM=INTERCEPT time1 | SUBJECT(rid*stateid) COVTYPE(DIAG)
  /REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).

```

Fonte: O autor.

Com relação à interpretação do modelo 3, temos que:

- dentre os coeficientes estimados, *percentFinAid* não se mostrou estatisticamente significativa ao nível de 0,05 ($p\text{-valor} = 0,198$);
- com relação ao intercepto (β_{000}), temos que, no *status* inicial ($t = 0$), a proporção de graduação por parte das instituições foi da ordem de 0,411 (41,1%), para uma instituição com uma porcentagem média (0) de estudantes recebendo ajuda financeira (*percentFinAid*) e com nível médio (0) de matrículas (*tuition*).

Figura 5.6 – Tela HWLM da saída do modelo 3

Fonte: O autor.

- O coeficiente do nível das matrículas (*tuition*) apresentou sinal positivo, indicando uma relação positiva com as proporções de graduação. Nesse caso, um aumento em um desvio padrão elevaria em, aproximadamente, 5,5% na taxa de graduação inicial.

Com relação aos efeitos aleatórios no nível dos estados (nível 3), temos que:

- após ajustar pelas covariáveis, a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de que não existe, no nível dos estados, variação entre as proporções de graduação no status inicial do estudo (intercepto) foi rejeitada (ao nível de 0,05), sugerindo uma variação da ordem de 0,249% (0,00249, $p\text{-valor} < 0,001$);
- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível dos estados também foi rejeitada, sugerindo uma taxa crescimento nas proporções de graduação entre os estados em torno de 0,03% (0,0003, $p\text{-valor} < 0,001$).

Com relação aos efeitos aleatórios no nível das instituições (nível 2), temos que:

- a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de inexistência de variação no *status* inicial entre as proporções de graduação no nível das instituições foi rejeitada (ao nível de 0,05), sugerindo uma variação de, aproximadamente, 1,657% (0,01657, $p\text{-valor} < 0,001$) entre as instituições.
- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível das instituições também foi rejeitada, sugerindo que as proporções de graduação entre instituições cresceram em torno de 0,78% (0,00775, $p\text{-valor} < 0,001$).

5.3.4 Modelo 4: Explorando Diferenças nas Trajetórias de Crescimento entre Instituições

Uma vez que foram constatadas evidências de variações no nível das instituições, duas novas variáveis (padronizadas) serão introduzidas em cada um dos parâmetros (ψ_0 e ψ_1), em um nível hierárquico mais elevado (nível 2): i) a seletividade no processo de admissão dos estudantes, *mathselect*, e ii) a porcentagem de professores em tempo integral, *percentFTfaculty*.

Ademais, no nível 3, serão também introduzidas duas outras variáveis (padronizadas): i) medida da porcentagem que as famílias pagam pela educação superior no estado, *avefamilyshare*, e ii) o percentual médio de calouros do próprio estado, *averetention*, que, de acordo com Hecker, Thomas e Tabata (2014, p. 264), também funciona como uma *proxy* da qualidade do sistema de educação pública superior do estado.

Dessa maneira, a forma do modelo 4 assume a forma:

$$y_{mti} = \psi_{0ti} + \psi_{1ti}time1_{mti} + \psi_{2ti}percentFinAid_{mti} + \psi_{3ti}tuition_{mti} + \varepsilon_{mti}. \quad (5.14)$$

No nível 2, teremos ψ_{0ti} e ψ_{1ti} dados por

$$\psi_{0ti} = \pi_{00i} + \pi_{01i}mathselect_{ti} + \pi_{02i}percentFTfaculty_{ti} + e_{0ti} \quad (5.15)$$

e

$$\psi_{1ti} = \pi_{10i} + \pi_{11i}mathselect_{ti} + \pi_{12i}percentFTfaculty_{ti} + e_{1ti}, \quad (5.16)$$

sendo ainda que, no nível 3, os interceptos do nível 2 serão dados por

$$\pi_{00i} = \beta_{000} + \beta_{001}avefamilyshare_i + \beta_{002}averetention_i + r_{00i} \quad (5.17)$$

e

$$\pi_{10i} = \beta_{100} + r_{10i}. \quad (5.18)$$

É importante ressaltar que o parâmetro de crescimento, ψ_{1ti} , passa a exibir duas interações de nível cruzado no nível das instituições.

As estimativas do modelo 4 encontram-se na Tabela 5.12.

Tabela 5.12 – Saídas das estimativas para o modelo 4 (continua)

<i>Efeito</i>		<i>Software</i>		
		HLM 7	R	SPSS
<i>Fixo</i>	Intercepto (β_{000})	0,481990	0,4774159	0,475
	Intercepto (β_{002}) <i>averetention</i>	0,049284	0,0495293	0,046
	Intercepto (β_{001}) <i>avefamilyshare</i>	0,018952	0,0194683	0,02
	Tabela 5.12 – Saídas das estimativas para o modelo 4 (conclusão)			
<i>Efeito</i>		<i>Software</i>		
		HLM 7	R	SPSS
<i>Aleatório</i>	Intercepto (π_{01i}) <i>mathselect</i>	0,092860	0,0934735	0,094
	Intercepto (π_{02i}) <i>percentFTfaculty</i>	0,037106	0,0369253	0,037
	time1 (β_{100}) <i>intercepto</i>	0,017661	0,0187478	0,019
	time1 (π_{11i}) <i>mathselect</i>	0,006005	0,0060438	0,006
	time1 (π_{12i}) <i>percentFTfaculty</i>	0,018600	0,0191958	0,019
	percentFinAid (ψ_{2ti})	0,001403	0,0017118	0,002
	tuition (ψ_{3ti})	0,049255	0,0448177	0,043
	Erro – nível 1 (ε_{mti})	0,00231	0,002655	0,003
	Erro – nível 2 (e_{0ti}) (intercepto)	0,00924	0,008635	0,007
	Erro – nível 2 (e_{1ti}) (time1)	0,00734	0,005436	0,004
	Erro – nível 3 (r_{00i}) (intercepto)	0,00067	0,0007568	0,001
	Erro – nível 3 (r_{10i}) (time1)	0,00031	0,0003102	0,000

Fonte: O autor.

Para a obtenção das saídas do software R, o comando que deverá ser executado encontra-se na Tabela 5.13.

No caso específico da execução do comando `fit4`, devemos chamar a atenção para o fato de haver 20 dados faltantes na variável *aveRetention* e 20 dados faltantes em

aveFamilyshare. Por esse motivo, foi necessário acrescentar o argumento `na.action = na.omit` no comando `lme`.

Além disso, pela Tabela 5.13, é possível notar que, à medida que um modelo linear hierárquico passa a assumir formas mais elaboradas, conhecimentos mais específicos de programação em R passam a ser exigidos por parte do pesquisador.

Tabela 5.13 – Comando executado pelo R para o modelo 4

```
fit4 <- lme(gradproportion ~ aveFamilyshare + aveRetention + percentFinAid + tuition + mathselect*time1 + percentFTfaculty*time1, random = ~ time1 | stateid/rid, correlation = corAR1(form = ~ 1 | stateid/rid), na.action=na.omit)
```

Fonte: O autor.

A saída do modelo 4 no R será dada pelo comando `summary(fit4)`, com as estimativas dos componentes de variância também na forma de desvio padrão.

Com relação ao SPSS, os comandos para o modelo 4 estão dispostos na Tabela 5.14.

Com relação ao HLM 7, a visualização do modelo 4 encontra-se na Figura 7. A simples substituição das equações de nível 3 nas equações de nível 2 e estas nas equações de nível 1 também produzirá a expressão descrita na parte inferior da Figura 7 (*Mixed Model*).

Tabela 5.14 – Comandos executados pelo SPSS para o modelo 4

```
MIXED gradproportion WITH mathselect percentFTfaculty aveRetention
aveFamilyshare time1
percentFinAid tuition
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
/FIXED=mathselect percentFTfaculty aveRetention aveFamilyshare time1
percentFinAid tuition
time1*mathselect time1*percentFTfaculty | SSTYPE(3)
/METHOD=REML
/PRINT=G SOLUTION TESTCOV
/RANDOM=INTERCEPT time1 | SUBJECT(stateid) COVTYPE(DIAG)
/RANDOM=INTERCEPT time1 | SUBJECT(rid*stateid) COVTYPE(DIAG)
/REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).
/RANDOM=INTERCEPT time1 | SUBJECT(stateid) COVTYPE(DIAG)
/RANDOM=INTERCEPT time1 | SUBJECT(rid*stateid) COVTYPE(DIAG)
/REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).
```

Fonte: O autor.

Ainda sobre a tela HWLM apresentada na Figura 7, é possível perceber que, mesmo se tratando de um modelo mais complexo, o software HLM oferece todos os seus aspectos relevantes do modelo desejado em uma única tela e de modo permanente. Essa conjunção de recursos acaba por tornar o próprio processo de elaboração de modelos mais produtivo ao oferecer contato visual com a forma do modelo desejado e acompanhamento permanente do processo de elaboração do mesmo.

Com relação à interpretação do modelo 4, temos, de um modo geral, que:

- dentre todos os parâmetros estimados do nível 1, o coeficiente ψ_{2ti} (0,0014, $p\text{-valor} = 0,207$) associado à variável *percentFinAid* não se mostrou estatisticamente significativo ao nível de 0,05;
- após ajuste do modelo levando em consideração as novas variáveis institucionais do nível 2 (*mathselect* e *percentFTfaculty*), a proporção de graduação por parte das instituições foi da ordem de 0,48199 (48,20%), para uma instituição com uma porcentagem média (0) de estudantes recebendo ajuda financeira (*percentFinAid*) e com nível médio (0) de matrículas (*tuition*).
- No nível 2 (entre instituições), *mathselect* e *percentFTfaculty* (em ψ_{0ti}) mostraram-se estatisticamente significantes, ao nível de 0,05. Mas, ao analisar ψ_{1ti} como função dessas mesmas variáveis (*mathselect* e *percentFTfaculty*), o coeficiente de interação *time1*mathselect*, π_{11i} (0,006, $p\text{-valor} = 0,213$), não se mostrou estatisticamente significativo.
- As variáveis *averetention* e *avefamilyshare* (efeitos fixos no nível 3) mostraram-se estatisticamente significantes (ao nível de 0,05) para explicar as proporções de graduação.
- O coeficiente associado ao termo de interação *time1*percentFTfaculty* mostrou-se positivamente associado com a proporção de graduação (1,86%, $p\text{-valor} < 0,001$), de forma que, considerando tudo mais constante, um aumento de um desvio padrão em *percentFTfaculty* tenderia a elevar em 1,86% a taxa de crescimento das proporções de graduação no período considerado, em comparação com instituições com valores médios (0) na variável *percentFTfaculty*.

Com relação aos efeitos aleatórios no nível dos estados (nível 3), temos que:

- após ajustar pelas novas variáveis (*averetention* e *avefamilyshare*), a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de que não existe, no nível dos estados, variação entre as proporções de graduação no *status* inicial do estudo

(intercepto) foi rejeitada (ao nível de 0,05), sugerindo uma variação da ordem de 0,067% (0,00067, p -valor < 0,001);

Figura 5.7 – Tela HWLM da saída do modelo 4

Fonte: O autor.

- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível dos estados também foi rejeitada, sugerindo uma taxa de crescimento nas proporções de graduação entre os estados em torno de 0,031% (0,00031, p -valor = 0,011).

Com relação aos efeitos aleatórios no nível das instituições (nível 2), temos que:

- a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de inexistência de variação no *status* inicial entre as proporções de graduação no nível das instituições foi rejeitada (ao nível de 0,05), sugerindo uma variação de, aproximadamente, 0,924% (0,00924, p -valor < 0,001) entre as instituições;
- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível das instituições também foi rejeitada, sugerindo

que as proporções de graduação entre instituições cresceram em torno de 0,734% (0,00734, p -valor < 0,001).

5.3.5 Modelo 5: Examinar as Taxas de Crescimento no Nível 3

No modelo 5, o coeficiente de *time1*, ψ_{1ti} , também será modelado por variáveis no nível 3, isto é, será avaliada a possibilidade de variáveis do nível 3 influenciarem as taxas de crescimento das proporções de graduação das instituições, ao longo do tempo. Dessa forma, a única modificação para o modelo 5 ocorrerá no nível 3.

A forma do modelo 5 será dada por

$$y_{mti} = \psi_{0ti} + \psi_{1ti}time1_{mti} + \psi_{2ti}percentFinAid_{mti} + \psi_{3ti}tuition_{mti} + \varepsilon_{mti}. \quad (5.19)$$

No nível 2, teremos

$$\psi_{0ti} = \pi_{00i} + \pi_{01i}mathselect_{ti} + \pi_{02i}percentFTfaculty_{ti} + e_{0ti} \quad (5.20)$$

e

$$\psi_{1ti} = \pi_{10i} + \pi_{11i}mathselect_{ti} + \pi_{12i}percentFTfaculty_{ti} + e_{1ti}, \quad (5.21)$$

sendo que, no nível 3, os interceptos do nível 2 assumirão a forma

$$\pi_{00i} = \beta_{000} + \beta_{001}avefamilyshare_i + \beta_{002}averetention_i + r_{00i} \quad (5.22)$$

e

$$\pi_{10i} = \beta_{100} + \beta_{101}avefamilyshare_i + \beta_{102}averetention_i + r_{10i} \quad (5.23)$$

As estimativas do modelo 5 encontram-se na Tabela 5.15.

Tabela 5.15 – Saídas das estimativas para o modelo 5

Efeito		Software		
		HLM 7	R	SPSS
Fixo	Intercepto (β_{000})	0,481656	0,476999	0,475
	Intercepto (β_{001}) <i>averetention</i>	0,042787	0,043257	0,044
	Intercepto (β_{002}) <i>avefamilyshare</i>	0,020217	0,02039	0,02
	Intercepto (π_{01i}) <i>mathselect</i>	0,093695	0,09428	0,094
	Intercepto (π_{02i}) <i>percentFTfaculty</i>	0,036726	0,0365148	0,037
	time1 (β_{100}) <i>intercepto (N3)</i>	0,018007	0,01921	0,020
	time1 (β_{102}) <i>averetention</i>	0,012157	0,011611	0,011
	time1 (β_{101}) <i>avefamilyshare</i>	-0,0028	-0,0019787	-0,002
	time1 (π_{11i}) <i>mathselect</i>	0,004467	0,0045681	0,005
	time1 (π_{12i}) <i>percentFTfaculty</i>	0,019198	0,0198902	0,020
	percentFinAid (ψ_{2ti})	0,001425	0,0017302	0,002
	tuition (ψ_{3ti})	0,049113	0,0446365	0,043

	Erro – nível 1 (ε_{mti})	0,00231	0,002654	0,003
	Erro – nível 2 (e_{0ti}) (intercepto)	0,00921	0,0086137	0,007
<i>Aleatório</i>	Erro – nível 2 (e_{1ti}) (time1)	0,00734	0,005429	0,004
	Erro – nível 3 (r_{00i}) (intercepto)	0,00069	0,000797	0,001
	Erro – nível 3 (r_{10i}) (time1)	0,00017	0,0002291	0,000

Fonte: O autor.

Para a obtenção das saídas do software R, o comando a ser executado encontra-se na Tabela 5.16.

Tabela 5.16 – Comando executado pelo R para o modelo 4

```
fit5 <- lme(gradproportion ~ percentFinAid + tuition + aveFamili
yshare*time1 + aveRetention*time1 + mathselect*time1 + percentF
Tfaculty*time1, random = ~ time1 | stateid/rid, correlation = c
orAR1(form = ~ 1 | stateid/rid), na.action = na.omit)
```

Fonte: O autor.

O objeto `fit5`, por incluir também as variáveis do nível 3 *aveRetention* e *aveFamilyshare*, deve receber o mesmo tratamento do objeto `fit4`, no que se refere ao acréscimo do argumento `na.action = na.omit`, uma vez que, conforme subseção anterior, cada uma dessas variáveis apresenta 20 dados faltantes.

A saída do modelo 5 no R será dada pelo comando `summary(fit5)`, com as estimativas dos componentes de variância também sob a forma de desvio padrão.

Com relação ao SPSS, os comandos do log estão dispostos na Tabela 5.17.

Tabela 5.17– Comandos executados pelo SPSS para o modelo 5

```
MIXED gradproportion WITH mathselect percentFTfaculty aveRetention
aveFamilyshare time1
    percentFinAid tuition
    /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
SINGULAR(0.0000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
PCONVERGE(0.000001, ABSOLUTE)
    /FIXED=mathselect percentFTfaculty aveRetention aveFamilyshare time1
percentFinAid tuition
    time1*aveFamilyshare      time1*aveRetention      time1*mathselect
time1*percentFTfaculty | SSTYPE(3)
    /METHOD=REML
    /PRINT=G SOLUTION TESTCOV
    /RANDOM=INTERCEPT time1 | SUBJECT(stateid) COVTYPE(DIAG)
    /RANDOM=INTERCEPT time1 | SUBJECT(rid*stateid) COVTYPE(DIAG)
    /REPEATED=time | SUBJECT(rid*stateid) COVTYPE(AR1).
```

Fonte: O autor.

Com relação ao HLM 7, a visualização do modelo 5 encontra-se na Figura 8. As substituições das equações de nível 3 nas equações de nível 2 e estas nas equações de nível 1 também produzem a expressão descrita na parte inferior da Figura 7 (*Mixed Model*).

Com relação à interpretação, chamaremos atenção para os resultados dos efeitos recém-introduzidos no modelo 5, uma vez que, com relação ao modelo 4, os resultados mostraram-se consistentes.

- O coeficiente β_{102} , associado ao termo de interação do nível 3 *time1*avefamilyshare* (-0,0028, *p-valor* = 0,533), não se mostrou estatisticamente significativo, ao nível de 0,05. Isso significa que a variável de nível 3 *avefamilyshare* não exerceu influência sobre o coeficiente ψ_{1ti} das taxas de crescimento da graduação.
- Contudo, o coeficiente β_{101} (0,012157, *p-valor* = 0,006), associado à variável do nível 3 *averetention*, mostrou-se estatisticamente significativo ao nível de 0,05, sugerindo uma relação positiva com o coeficiente das taxas de crescimento da graduação ψ_{1ti} . Isso significa que o aumento de um desvio padrão na proporção da retenção dos alunos, em nível estadual, tenderia a elevar em, aproximadamente, 1,22% a proporção de graduados no período considerado, mantendo todas as demais variáveis constantes.

Com relação aos efeitos aleatórios no nível dos estados (nível 3), temos que:

- a hipótese nula de que não existe, no nível dos estados, variação entre as proporções de graduação no status inicial do estudo (intercepto) foi rejeitada (ao nível de 0,05), sugerindo uma variação da ordem de 0,069% (0,00069, *p-valor* < 0,001).
- Para a variância do coeficiente *time1*, após ajustar pelas novas variáveis (*averetention* e *avefamilyshare*) em π_{10i} , a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível dos estados também foi rejeitada ao nível de 0,05, sugerindo uma taxa crescimento nas proporções de graduação entre os estados em torno de 0,017% (0,00017, *p-valor* = 0,045).

Com relação aos efeitos aleatórios no nível das instituições (nível 2), temos que:

- a hipótese nula de inexistência de variabilidade no intercepto, isto é, a hipótese de inexistência de variação no *status* inicial entre as proporções de graduação no nível das instituições foi rejeitada (ao nível de 0,05), sugerindo uma variação de, aproximadamente, 0,921% (0,00921, *p-valor* < 0,001) entre as instituições;

Figura 5.8 – Tela HWLM da saída do modelo 5

Fonte: O autor.

- para a variância do coeficiente *time1*, a hipótese nula de inexistência de variação nos parâmetros de crescimentos no nível das instituições também foi rejeitada, sugerindo que as proporções de graduação entre instituições cresceram em torno de 0,734% (0,00734, *p*-valor < 0,001), praticamente sem mudanças quando comparada com a estimativa do modelo 4.

Dessa maneira, o software HLM 7 produziu três resultados que merecem destaque:

- 1) as estimativas dos parâmetros estiveram bem próximas daquelas produzidas pelos outros dois softwares (SPSS e R);
- 2) forneceu estimativas de *p*-valor de todos os componentes de variância (níveis 2 e 3), fato que também ocorreu com o SPSS, mas não com a biblioteca *nlme* do software R;
- 3) em um contexto de análise de dados hierárquicos, caso um pesquisador decida trabalhar com a estratégia *three-level growth formulation*, o software HLM 7, por meio de simples atualizações realizadas na tela HWLM, fornece as estimativas de modelos hierárquicos

de forma simples e intuitiva, sem a necessidade de, por um lado, passar por várias telas (como é o caso do SPSS) e, por outro, de conhecimentos específicos de programação (como é o caso do R).

Com relação à segunda consideração, é necessário reconhecer a existência de um debate na comunidade científica sobre certos aspectos técnicos do cálculo do *p-valor* para as estimativas da variância dos termos aleatórios (componentes de variância).

Pinheiros e Bates (2000, p. 84) lembram que Stram e Lee (1994), fazendo uso de resultados de Self e Liang (1987), alegam que os resultados dos testes de efeitos aleatórios de modelos aninhados (p. ex., um modelo com intercepto aleatório e um outro com intercepto e coeficiente aleatórios) podem ser *conservativos*, no sentido de que o *p-valor* obtido em um teste da razão de verossimilhança é maior do que deveria ser.

A explicação de Stram e Lee (1994), de acordo com Pinheiro e Bates (2000), reside no fato de que a mudança para um modelo mais específico (p. ex., intercepto e coeficiente aleatórios), a partir de um modelo mais geral (p. ex., intercepto aleatório), implica em estabelecer, nessa passagem, a variância de um componente aleatório (ou mais, se for o caso) igual a zero, limite do espaço paramétrico.

De acordo, ainda, com Pinheiros e Bates (2000, p. 86), Stram e Lee (1994) sugeriram um ajuste baseado em uma mistura de distribuições qui-quadrado que teria produzido resultados interessantes, mas, que, não constituiria uma solução geral. Em razão disso, recomendam cuidado com o uso de *p-valor* especificamente para avaliações de efeitos aleatórios em modelos de efeitos mistos (Pinheiros e Bates, 2000, p. 87).

5.4 Apêndice

HLM 7 – Saída do modelo nulo

05/11/2017

hlm3.html

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
 Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2010
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM3.EXE (7.01.21202.1001)
 Date: 5 November 2017, Sunday
 Time: 0:48:43

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = chap6_new.mdm
 The command file for this run = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software
 HLM7\HLM7_chap6_null_model.hlm
 Output file name = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\hlm3.html
 The maximum number of level-1 units = 7022
 The maximum number of level-2 units = 647
 The maximum number of level-3 units = 50
 The maximum number of iterations = 100
 Method of estimation: full maximum likelihood
 The outcome variable is GRADPROP

Summary of the model specified

Level-1 Model

$$\text{GRADPROP}_{mti} = \psi_{0ti} + \varepsilon_{mti}$$

Level-2 Model

$$\psi_{0ti} = \pi_{00i} + e_{0ti}$$

Level-3 Model

$$\pi_{00i} = \beta_{000} + r_{00i}$$

Mixed Model

$$\text{GRADPROP}_{mti} = \beta_{000} + e_{0ti} + r_{00i} + \varepsilon_{mti}$$

For starting values, data from 7022 level-1 and 647 level-2 records were used

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

1/3

HLM 7 – Saída do modelo nulo (continuação)

05/11/2017

hlm3.html

Final Results - Iteration 6**Iterations stopped due to small change in likelihood function**

$$\sigma^2_e = 0.00340$$

Standard error of $\sigma^2_e = 0.00006$

$$\sigma^2_e$$

INTRCPT1, ψ_0 0.01703Standard error of σ^2_e INTRCPT1, ψ_0 0.00100

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.981

$$\tau_\pi$$

INTRCPT1

INTRCPT2, π_{00}

0.00319

Standard error of τ_π

INTRCPT1

INTRCPT2, π_{00}

0.00099

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, π_{00}	0.628

The value of the log-likelihood function at iteration 6 = 8.670970E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.431584	0.010075	42.838	49	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.431584	0.010073	42.846	49	<0.001

Final estimation of level-1 and level-2 variance components

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

2/3

HLM 7 – Saída do modelo nulo (continuação)

05/11/2017

hlm3.html

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1, e_{θ}	0.13049	0.01703	597	33269.65268	<0.001
level-1, ε	0.05830	0.00340			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1/INTRCPT2, $r_{\theta\theta}$	0.05647	0.00319	49	169.57049	<0.001

Statistics for the current model

Deviance = -17341.939481

Number of estimated parameters = 4

HLM 7 – Saída do modelo 1

06/11/2017

hlm3.html

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
 Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2010
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM3.EXE (7.01.21202.1001)
 Date: 6 November 2017, Monday
 Time: 1:24:21

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = chap6_new.mdm
 The command file for this run = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software
 HLM7\HLM7_chap6_model1a.hlm
 Output file name = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\hlm3.html
 The maximum number of level-1 units = 7022
 The maximum number of level-2 units = 647
 The maximum number of level-3 units = 50
 The maximum number of iterations = 100
 Method of estimation: full maximum likelihood
 The outcome variable is GRADPROP

Summary of the model specified

Level-1 Model

$$\text{GRADPROP}_{mi} = \psi_{0i} + \psi_{1i} * (\text{TIME1}_{mi}) + \varepsilon_{mi}$$

Level-2 Model

$$\begin{aligned}\psi_{0i} &= \pi_{00i} + e_{0i} \\ \psi_{1i} &= \pi_{10i} + e_{1i}\end{aligned}$$

Level-3 Model

$$\begin{aligned}\pi_{00i} &= \beta_{000} + r_{00i} \\ \pi_{10i} &= \beta_{100} + r_{10i}\end{aligned}$$

Mixed Model

HLM 7 – Saída do modelo 1 (continuação)

06/11/2017

hlm3.html

$$GRADPROP_{mti} = \beta_{000} + \beta_{100} * TIME1_{mti} + e_{0ti} + e_{1ti} * TIME1_{mti} + r_{00i} + r_{10i} * TIME1_{mti} + \varepsilon_{mti}$$

For starting values, data from 7018 level-1 and 643 level-2 records were used

Final Results - Iteration 17

Iterations stopped due to small change in likelihood function

$$\sigma^2_{\varepsilon} = 0.00234$$

Standard error of $\sigma^2_{\varepsilon} = 0.00004$

$$\sigma^2_{\varepsilon}$$

INTRCPT1, ψ_0 0.01726 -0.00206

TIME1, ψ_1 -0.00206 0.00771

Standard errors of σ^2_{ε}

INTRCPT1, ψ_0 0.00104 0.00056

TIME1, ψ_1 0.00056 0.00057

σ^2_{ε} (as correlations)

INTRCPT1, ψ_0 1.000 -0.178

TIME1, ψ_1 -0.178 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.954
TIME1, ψ_1	0.777

Note: The reliability estimates reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

$$\tau_{\pi}$$

INTRCPT1 TIME1

INTRCPT2, π_{00} INTRCPT2, π_{10}

0.00282 0.00030

0.00030 0.00035

Standard errors of τ_{π}

INTRCPT1 TIME1

INTRCPT2, π_{00} INTRCPT2, π_{10}

0.00091 0.00032

0.00032 0.00023

τ_{π} (as correlations)

INTRCPT1/INTRCPT2, π_{00} 1.000 0.301

TIME1/INTRCPT2, π_{10} 0.301 1.000

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, π_{00}	0.594
TIME1/INTRCPT2, π_{10}	0.278

The value of the log-likelihood function at iteration 17 = 9.370654E+003

HLM 7 – Saída do modelo 1 (continuação)

06/11/2017

hlm3.html

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.411065	0.009710	42.332	49	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.040737	0.004984	8.173	49	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.411065	0.009710	42.335	49	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.040737	0.004978	8.183	49	<0.001

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, e_0	0.13139	0.01726	593	14280.22215	<0.001
TIME1 slope, e_1	0.08783	0.00771	593	2668.53762	<0.001
level-1, ε	0.04839	0.00234			

Note: The chi-square statistics reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2, r_{00}	0.05306	0.00282	49	151.13946	<0.001
TIME1/INTRCPT2, r_{10}	0.01863	0.00035	49	80.60127	0.003

Statistics for the current model

Deviance = -18741.307425

Number of estimated parameters = 9

HLM 7 – Saída do modelo 2

05/11/2017

hlm3.html

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
 Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2010
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM3.EXE (7.01.21202.1001)
 Date: 5 November 2017, Sunday
 Time: 6:26:29

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = chap6_new.mdm
 The command file for this run = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software
 HLM7\HLM7_chap6_model2a.hlm
 Output file name = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\hlm3.html
 The maximum number of level-1 units = 7022
 The maximum number of level-2 units = 647
 The maximum number of level-3 units = 50
 The maximum number of iterations = 100
 Method of estimation: full maximum likelihood
 The outcome variable is GRADPROP

Summary of the model specified

Level-1 Model

$$\text{GRADPROP}_{mti} = \psi_{0ti} + \psi_{1ti} * (\text{TIME1}_{mti}) + \psi_{2ti} * (\text{PERCENTF}_{mti}) + \psi_{3ti} * (\text{TUITION}_{mti}) + \varepsilon_{mti}$$

Level-2 Model

$$\psi_{0ti} = \pi_{00i} + e_{0ti}$$

$$\psi_{1ti} = \pi_{10i} + e_{1ti}$$

$$\psi_{2ti} = \pi_{20i}$$

$$\psi_{3ti} = \pi_{30i}$$

Level-3 Model

$$\pi_{00i} = \beta_{000} + r_{00i}$$

$$\pi_{10i} = \beta_{100} + r_{10i}$$

$$\pi_{20i} = \beta_{200}$$

$$\pi_{30i} = \beta_{300}$$

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

1/4

HLM 7 – Saída do modelo 2 (continuação)

05/11/2017

hlm3.html

Mixed Model

$$GRADPROP_{mti} = \beta_{000} + \beta_{100} * TIME1_{mti} + \beta_{200} * PERCENTF_{mti} + \beta_{300} * TUITION_{mti} + e_{0ti} + e_{1ti} * TIME1_{mti} + r_{00i} + r_{10i} * TIME1_{mti} + \varepsilon_{mti}$$

For starting values, data from 7018 level-1 and 643 level-2 records were used

Final Results - Iteration 9**Iterations stopped due to small change in likelihood function**

$$\sigma^2_{\varepsilon} = 0.00231$$

Standard error of $\sigma^2_{\varepsilon} = 0.00004$

$$\sigma^2_{\varepsilon}$$

INTRCPT1, ψ_0 0.01657 -0.00256

TIME1, ψ_1 -0.00256 0.00775

Standard errors of σ^2_{ε}

INTRCPT1, ψ_0 0.00100 0.00056

TIME1, ψ_1 0.00056 0.00057

σ^2_{ε} (as correlations)

INTRCPT1, ψ_0 1.000 -0.226

TIME1, ψ_1 -0.226 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.953
TIME1, ψ_1	0.781

Note: The reliability estimates reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

$$\tau_{\pi}$$

INTRCPT1 TIME1

INTRCPT2, π_{00} INTRCPT2, π_{10}

0.00249 0.00021

0.00021 0.00030

Standard errors of τ_{π}

INTRCPT1 TIME1

INTRCPT2, π_{00} INTRCPT2, π_{10}

0.00083 0.00030

0.00030 0.00021

τ_{π} (as correlations)

INTRCPT1/INTRCPT2, π_{00} 1.000 0.240

TIME1/INTRCPT2, π_{10} 0.240 1.000

Random level-2 coefficient	Reliability estimate
----------------------------	----------------------

HLM 7 – Saída do modelo 2 (continuação)

05/11/2017

hlm3.html

INTRCPT1/INTRCPT2, π_{00} 0.576TIME1/INTRCPT2, π_{10} 0.251

The value of the log-likelihood function at iteration 9 = 9.432350E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.475789	0.010966	43.388	49	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.014657	0.005396	2.716	49	0.009
For PERCENTF slope, ψ_2					
For INTRCPT2, π_{20}					
INTRCPT3, β_{200}	0.001444	0.000860	1.679	5626	0.093
For TUITION slope, ψ_3					
For INTRCPT2, π_{30}					
INTRCPT3, β_{300}	0.054975	0.004999	10.998	5626	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.475789	0.014981	31.760	49	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.014657	0.006183	2.371	49	0.022
For PERCENTF slope, ψ_2					
For INTRCPT2, π_{20}					
INTRCPT3, β_{200}	0.001444	0.001120	1.289	5626	0.198
For TUITION slope, ψ_3					
For INTRCPT2, π_{30}					
INTRCPT3, β_{300}	0.054975	0.008806	6.243	5626	<0.001

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, e_0	0.12871	0.01657	593	13956.18048	<0.001
TIME1 slope, e_1	0.08806	0.00775	593	2733.53252	<0.001
level-1, ε	0.04806	0.00231			

Note: The chi-square statistics reported above are based on only 643 of 647

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

3/4

HLM 7 – Saída do modelo 2 (continuação)

05/11/2017

hlm3.html

units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	<i>d.f.</i>	χ^2	<i>p</i> -value
INTRCPT1/INTRCPT2, r_{00}	0.04994	0.00249	49	141.71551	<0.001
TIME1/INTRCPT2, r_{10}	0.01731	0.00030	49	73.70734	0.013

Statistics for the current model

Deviance = -18864.699972

Number of estimated parameters = 11

HLM 7 – Saída do modelo 3

05/11/2017

hlm3.html

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
 Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2010
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM3.EXE (7.01.21202.1001)
 Date: 5 November 2017, Sunday
 Time: 7:35:49

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = chap6_new.mdm
 The command file for this run = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software
 HLM7\HLM7_chap6_model3a.hlm
 Output file name = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\hlm3.html
 The maximum number of level-1 units = 7022
 The maximum number of level-2 units = 647
 The maximum number of level-3 units = 50
 The maximum number of iterations = 100
 Method of estimation: full maximum likelihood
 The outcome variable is GRADPROP

Summary of the model specified

Level-1 Model

$$\text{GRADPROP}_{mti} = \psi_{0ti} + \psi_{1ti} * (\text{TIME1}_{mti}) + \psi_{2ti} * (\text{PERCENTF}_{mti}) + \psi_{3ti} * (\text{TUITION}_{mti}) + \varepsilon_{mti}$$

Level-2 Model

$$\begin{aligned}\psi_{0ti} &= \pi_{00i} + \pi_{01i} * (\text{MATHSELE}_{ti}) + \pi_{02i} * (\text{PERFTFAC}_{ti}) + e_{0ti} \\ \psi_{1ti} &= \pi_{10i} + \pi_{11i} * (\text{MATHSELE}_{ti}) + \pi_{12i} * (\text{PERFTFAC}_{ti}) + e_{1ti} \\ \psi_{2ti} &= \pi_{20i} \\ \psi_{3ti} &= \pi_{30i}\end{aligned}$$

Level-3 Model

$$\begin{aligned}\pi_{00i} &= \beta_{000} + \beta_{001}(\text{AVERETEN}_i) + \beta_{002}(\text{AVEFAMIL}_i) + r_{00i} \\ \pi_{01i} &= \beta_{010} \\ \pi_{02i} &= \beta_{020} \\ \pi_{10i} &= \beta_{100} + r_{10i}\end{aligned}$$

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

1/4

HLM 7 – Saída do modelo 3 (continuação)

05/11/2017

hlm3.html

$$\pi_{11i} = \beta_{110}$$

$$\pi_{12i} = \beta_{120}$$

$$\pi_{20i} = \beta_{200}$$

$$\pi_{30i} = \beta_{300}$$

Mixed Model

$$\begin{aligned} GRADPROP_{mti} = & \beta_{000} + \beta_{001} * AVERETEN_i + \beta_{002} * AVEFAMIL_i + \beta_{010} * MATHSELE_{ti} \\ & + \beta_{020} * PERFTFAC_{ti} + \beta_{100} * TIME1_{mti} + \beta_{110} * TIME1_{mti} * MATHSELE_{ti} + \beta_{120} * TIME1_{mti} * PERFTFAC_{ti} \\ & + \beta_{200} * PERCENTF_{mti} + \beta_{300} * TUITION_{mti} \\ & + e_{0ti} + e_{1ti} * TIME1_{mti} + r_{00i} + r_{10i} * TIME1_{mti} + e_{mti} \end{aligned}$$

For starting values, data from 7018 level-1 and 643 level-2 records were used

Final Results - Iteration 17**Iterations stopped due to small change in likelihood function**

$$\sigma^2_e = 0.00231$$

Standard error of $\sigma^2_e = 0.00004$

$$\sigma^2_e$$

$$\text{INTRCPT1}, \psi_0 \quad 0.00924 \quad -0.00382$$

$$\text{TIME1}, \psi_1 \quad -0.00382 \quad 0.00734$$

Standard errors of σ^2_e

$$\text{INTRCPT1}, \psi_0 \quad 0.00058 \quad 0.00045$$

$$\text{TIME1}, \psi_1 \quad 0.00045 \quad 0.00055$$

σ^2_e (as correlations)

$$\text{INTRCPT1}, \psi_0 \quad 1.000 \quad -0.463$$

$$\text{TIME1}, \psi_1 \quad -0.463 \quad 1.000$$

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.921
TIME1, ψ_1	0.771

Note: The reliability estimates reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

$$\tau_\pi$$

$$\text{INTRCPT1} \quad \text{TIME1}$$

$$\text{INTRCPT2}, \pi_{00} \quad \text{INTRCPT2}, \pi_{10}$$

$$0.00067 \quad -0.00017$$

$$-0.00017 \quad 0.00031$$

Standard errors of τ_π

$$\text{INTRCPT1} \quad \text{TIME1}$$

$$\text{INTRCPT2}, \pi_{00} \quad \text{INTRCPT2}, \pi_{10}$$

HLM 7 – Saída do modelo 3 (continuação)

05/11/2017

hlm3.html

0.00031 0.00020
 0.00020 0.00021

 τ_π (as correlations)INTRCPT1/INTRCPT2, π_{00} 1.000 -0.385TIME1/INTRCPT2, π_{10} -0.385 1.000

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, π_{00}	0.406
TIME1/INTRCPT2, π_{10}	0.264

The value of the log-likelihood function at iteration 17 = 9.701785E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.481990	0.007977	60.422	47	<0.001
AVERETEN, β_{001}	0.049284	0.004996	9.864	47	<0.001
AVEFAMIL, β_{002}	0.018952	0.005030	3.768	47	<0.001
For MATHSELE, π_{01}					
INTRCPT3, β_{010}	0.092860	0.005549	16.735	543	<0.001
For PERFTFAC, π_{02}					
INTRCPT3, β_{020}	0.037106	0.004343	8.544	543	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.017661	0.005375	3.286	49	0.002
For MATHSELE, π_{11}					
INTRCPT3, β_{110}	0.006005	0.005324	1.128	543	0.260
For PERFTFAC, π_{12}					
INTRCPT3, β_{120}	0.018600	0.004123	4.511	543	<0.001
For PERCENTF slope, ψ_2					
For INTRCPT2, π_{20}					
INTRCPT3, β_{200}	0.001403	0.000855	1.641	5626	0.101
For TUITION slope, ψ_3					
For INTRCPT2, π_{30}					
INTRCPT3, β_{300}	0.049255	0.004762	10.344	5626	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.481990	0.011001	43.813	47	<0.001
AVERETEN, β_{001}	0.049284	0.005108	9.649	47	<0.001
AVEFAMIL, β_{002}	0.018952	0.004817	3.934	47	<0.001

HLM 7 – Saída do modelo 3 (continuação)

05/11/2017

hlm3.html

For MATHSELE, π_{01}						
INTRCPT3, β_{010}	0.092860	0.007890	11.769	543	<0.001	
For PERFTFAC, π_{02}						
INTRCPT3, β_{020}	0.037106	0.006052	6.131	543	<0.001	
For TIME1 slope, ψ_1						
For INTRCPT2, π_{10}						
INTRCPT3, β_{100}	0.017661	0.006327	2.791	49	0.007	
For MATHSELE, π_{11}						
INTRCPT3, β_{110}	0.006005	0.004815	1.247	543	0.213	
For PERFTFAC, π_{12}						
INTRCPT3, β_{120}	0.018600	0.005026	3.701	543	<0.001	
For PERCENTF slope, ψ_2						
For INTRCPT2, π_{20}						
INTRCPT3, β_{200}	0.001403	0.001111	1.263	5626	0.207	
For TUITION slope, ψ_3						
For INTRCPT2, π_{30}						
INTRCPT3, β_{300}	0.049255	0.007727	6.374	5626	<0.001	

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, e_0	0.09611	0.00924	591	7929.37469	<0.001
TIME1 slope, e_1	0.08569	0.00734	591	2620.93105	<0.001
level-1, ε	0.04805	0.00231			

Note: The chi-square statistics reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2, r_{00}	0.02579	0.00067	47	92.67866	<0.001
TIME1/INTRCPT2, r_{10}	0.01758	0.00031	49	74.62588	0.011

Statistics for the current model

Deviance = -19403.570662

Number of estimated parameters = 17

HLM 7 – Saída do modelo 4

05/11/2017

hlm3.html

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
 Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2010
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM3.EXE (7.01.21202.1001)
 Date: 5 November 2017, Sunday
 Time: 8:19:43

Specifications for this HLM3 run

Problem Title: no title

The data source for this run = chap6_new.mdm
 The command file for this run = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\HLM7_chap6_model4a.hlm
 Output file name = C:\Users\Koki\Desktop\Tese2\Capitulos\Cap 5 Software HLM7\hlm3.html
 The maximum number of level-1 units = 7022
 The maximum number of level-2 units = 647
 The maximum number of level-3 units = 50
 The maximum number of iterations = 100
 Method of estimation: full maximum likelihood
 The outcome variable is GRADPROP

Summary of the model specified

Level-1 Model

$$\text{GRADPROP}_{mti} = \psi_{0ti} + \psi_{1ti} * (\text{TIME1}_{mti}) + \psi_{2ti} * (\text{PERCENTF}_{mti}) + \psi_{3ti} * (\text{TUITION}_{mti}) + \varepsilon_{mti}$$

Level-2 Model

$$\begin{aligned}\psi_{0ti} &= \pi_{00i} + \pi_{01i} * (\text{MATHSELE}_{ti}) + \pi_{02i} * (\text{PERFTFAC}_{ti}) + e_{0ti} \\ \psi_{1ti} &= \pi_{10i} + \pi_{11i} * (\text{MATHSELE}_{ti}) + \pi_{12i} * (\text{PERFTFAC}_{ti}) + e_{1ti} \\ \psi_{2ti} &= \pi_{20i} \\ \psi_{3ti} &= \pi_{30i}\end{aligned}$$

Level-3 Model

$$\begin{aligned}\pi_{00i} &= \beta_{000} + \beta_{001}(\text{AVERETEN}_i) + \beta_{002}(\text{AVEFAMIL}_i) + r_{00i} \\ \pi_{01i} &= \beta_{010} \\ \pi_{02i} &= \beta_{020} \\ \pi_{10i} &= \beta_{100} + \beta_{101}(\text{AVERETEN}_i) + \beta_{102}(\text{AVEFAMIL}_i) + r_{10i}\end{aligned}$$

file:///C:/Users/Koki/Desktop/Tese2/Capitulos/Cap%205%20Software%20HLM7/hlm3.html

1/4

HLM 7 – Saída do modelo 4 (continuação)

05/11/2017

hlm3.html

$$\pi_{11i} = \beta_{110}$$

$$\pi_{12i} = \beta_{120}$$

$$\pi_{20i} = \beta_{200}$$

$$\pi_{30i} = \beta_{300}$$

Mixed Model

$$\begin{aligned} GRADPROP_{mti} = & \beta_{000} + \beta_{001} * AVERETEN_i + \beta_{002} * AVEFAMIL_i + \beta_{010} * MATHSELE_{ti} \\ & + \beta_{020} * PERFTFAC_{ti} + \beta_{100} * TIME1_{mti} + \beta_{101} * TIME1_{mti} * AVERETEN_i + \beta_{102} * TIME1_{mti} * AVEFAMIL_i \\ & + \beta_{110} * TIME1_{mti} * MATHSELE_{ti} + \beta_{120} * TIME1_{mti} * PERFTFAC_{ti} + \beta_{200} * PERCENTF_{mti} + \beta_{300} * TUITION_{mti} \\ & + e_{0ti} + e_{1ti} * TIME1_{mti} + r_{00i} + r_{10i} * TIME1_{mti} + e_{mti} \end{aligned}$$

For starting values, data from 7018 level-1 and 643 level-2 records were used

Final Results - Iteration 13**Iterations stopped due to small change in likelihood function**

$$\sigma^2_e = 0.00231$$

Standard error of $\sigma^2_e = 0.00004$

$$\sigma^2_e$$

$$\text{INTRCPT1}, \psi_0 \quad 0.00921 \quad -0.00379$$

$$\text{TIME1}, \psi_1 \quad -0.00379 \quad 0.00734$$

Standard errors of σ^2_e

$$\text{INTRCPT1}, \psi_0 \quad 0.00058 \quad 0.00044$$

$$\text{TIME1}, \psi_1 \quad 0.00044 \quad 0.00055$$

σ^2_e (as correlations)

$$\text{INTRCPT1}, \psi_0 \quad 1.000 \quad -0.461$$

$$\text{TIME1}, \psi_1 \quad -0.461 \quad 1.000$$

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.921
TIME1, ψ_1	0.771

Note: The reliability estimates reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

$$\tau_\pi$$

$$\text{INTRCPT1} \quad \text{TIME1}$$

$$\text{INTRCPT2}, \pi_{00} \quad \text{INTRCPT2}, \pi_{10}$$

$$0.00069 \quad -0.00016$$

$$-0.00016 \quad 0.00017$$

Standard errors of τ_π

$$\text{INTRCPT1} \quad \text{TIME1}$$

$$\text{INTRCPT2}, \pi_{00} \quad \text{INTRCPT2}, \pi_{10}$$

HLM 7 – Saída do modelo 4 (continuação)

05/11/2017

hlm3.html

0.00031 0.00018

0.00018 0.00017

 τ_π (as correlations)INTRCPT1/INTRCPT2, π_{00} 1.000 -0.478TIME1/INTRCPT2, π_{10} -0.478 1.000

Random level-2 coefficient	Reliability estimate
INTRCPT1/INTRCPT2, π_{00}	0.415
TIME1/INTRCPT2, π_{10}	0.172

The value of the log-likelihood function at iteration 13 = 9.705330E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.481656	0.008015	60.093	47	<0.001
AVERETEN, β_{001}	0.042787	0.005643	7.583	47	<0.001
AVEFAMIL, β_{002}	0.020217	0.005665	3.569	47	<0.001
For MATHSELE, π_{01}					
INTRCPT3, β_{010}	0.093695	0.005551	16.878	543	<0.001
For PERFTFAC, π_{02}					
INTRCPT3, β_{020}	0.036726	0.004349	8.445	543	<0.001
For TIME1 slope, ψ_1					
For INTRCPT2, π_{10}					
INTRCPT3, β_{100}	0.018007	0.005031	3.579	47	<0.001
AVERETEN, β_{101}	0.012157	0.004436	2.741	47	0.009
AVEFAMIL, β_{102}	-0.002800	0.004457	-0.628	47	0.533
For MATHSELE, π_{11}					
INTRCPT3, β_{110}	0.004467	0.005305	0.842	543	0.400
For PERFTFAC, π_{12}					
INTRCPT3, β_{120}	0.019198	0.004091	4.693	543	<0.001
For PERCENTF slope, ψ_2					
For INTRCPT2, π_{20}					
INTRCPT3, β_{200}	0.001425	0.000855	1.668	5626	0.095
For TUITION slope, ψ_3					
For INTRCPT2, π_{30}					
INTRCPT3, β_{300}	0.049113	0.004766	10.306	5626	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, ψ_0					
For INTRCPT2, π_{00}					
INTRCPT3, β_{000}	0.481656	0.011172	43.111	47	<0.001

HLM 7 – Saída do modelo 4 (continuação)

05/11/2017			hlm3.html			
	AVERETEN, β_{001}	0.042787	0.005389	7.940	47	<0.001
	AVEFAMIL, β_{002}	0.020217	0.005840	3.462	47	0.001
For MATHSELE, π_{01}						
	INTRCPT3, β_{010}	0.093695	0.008111	11.552	543	<0.001
For PERFTFAC, π_{02}						
	INTRCPT3, β_{020}	0.036726	0.006158	5.964	543	<0.001
For TIME1 slope, ψ_1						
For INTRCPT2, π_{10}						
	INTRCPT3, β_{100}	0.018007	0.006211	2.899	47	0.006
	AVERETEN, β_{101}	0.012157	0.004220	2.881	47	0.006
	AVEFAMIL, β_{102}	-0.002800	0.004988	-0.561	47	0.577
For MATHSELE, π_{11}						
	INTRCPT3, β_{110}	0.004467	0.005223	0.855	543	0.393
For PERFTFAC, π_{12}						
	INTRCPT3, β_{120}	0.019198	0.005153	3.725	543	<0.001
For PERCENTF slope, ψ_2						
For INTRCPT2, π_{20}						
	INTRCPT3, β_{200}	0.001425	0.001112	1.282	5626	0.200
For TUITION slope, ψ_3						
For INTRCPT2, π_{30}						
	INTRCPT3, β_{300}	0.049113	0.007665	6.408	5626	<0.001

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, e_0	0.09597	0.00921	591	7928.87404	<0.001
TIME1 slope, e_1	0.08566	0.00734	591	2620.38083	<0.001
level-1, ε	0.04805	0.00231			

Note: The chi-square statistics reported above are based on only 643 of 647 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2, r_{00}	0.02627	0.00069	47	93.35677	<0.001
TIME1/INTRCPT2, r_{10}	0.01304	0.00017	47	64.56914	0.045

Statistics for the current model

Deviance = -19410.660326

Number of estimated parameters = 19

6 ALGUMAS RELAÇÕES ENTRE OS MODELOS PA E SS

6.1 Introdução

Neste capítulo, serão abordadas as classes de modelos PA e SS. Apesar de inicialmente restritos a modelos de regressão logística, uma relação matemática exata entre os parâmetros dos modelos PA e SS foram estabelecidos. Posteriormente, essa relação matemática entre os parâmetros dos modelos PA e SS foi estendida para modelos de regressão de Poisson. Ao final deste capítulo, uma análise sobre algumas relações gerais de equivalência entre os parâmetros de modelos PA e SS será realizada.

6.2 Os modelos PA e SS – Regressão Logística

6.2.1 A Proposta de Zeger, Liang e Albert

Dois anos após a criação dos modelos GEE e quatro anos após a introdução dos modelos LMM, Zeger et al. (1988) discutiram sobre uma proposta baseada em duas estratégias para análise de dados longitudinais: i) modelos PA (*population-averaged*) e ii) modelos SS (*subject-specific*).

A classe de modelos PA compreenderia os modelos com foco central sobre a média marginal da variável resposta, tendo os parâmetros β uma interpretação em termos da média da população dos indivíduos (*Population Averaged*).

Os modelos SS estariam voltados para a modelagem da fonte de heterogeneidade das UEs, de modo que a interpretação dos parâmetros β , nesse caso, seria de ordem individual (*Subject Specific*).

Para ilustrar essa distinção na interpretação dos parâmetros, Hardin e Hilbe (2013, p. 60) apresentou um exemplo ilustrativo com a regressão logística longitudinal cuja variável resposta de interesse, Y_{it} , consistiria na condição de uma criança possuir, ou não, doença respiratória. Por simplificação, foi admitida apenas uma covariável, X_{it} , correspondendo ao *status* de fumante da mãe.

O modelo SS seria dado por

$$\text{logit } \mathbb{P}(Y_{it} = 1 | X_{it}, b_i) = \beta_0^{SS} + X_{it}\beta_1^{SS} + b_i, \quad (6.1)$$

de modo que a razão de chances, nesse caso, seria dada por

$$RC^{SS} = \frac{\mathbb{P}(Y_{it} = 1|X_{it} = 1, b_i)/\mathbb{P}(Y_{it} = 0|X_{it} = 1, b_i)}{\mathbb{P}(Y_{it} = 1|X_{it} = 0, b_i)/\mathbb{P}(Y_{it} = 0|X_{it} = 0, b_i)} = \exp(\beta^{SS}), \quad (6.2)$$

com $\exp(\beta^{SS})$ representando a razão da chance de uma dada criança possuir doença respiratória quando a mãe é fumante comparativamente à chance de a mesma criança ter uma doença respiratória quando a mãe não é fumante.

O modelo PA seria dado por

$$\text{logit } \mathbb{P}(Y_{it} = 1|X_{it}) = \beta_0^{PA} + X_{it}\beta_1^{PA}. \quad (6.3)$$

Nesse caso em particular, a razão de chances seria dada por

$$RC^{PA} = \frac{\mathbb{P}(Y_{it} = 1|X_{it} = 1)/\mathbb{P}(Y_{it} = 0|X_{it} = 1)}{\mathbb{P}(Y_{it} = 1|X_{it} = 0)/\mathbb{P}(Y_{it} = 0|X_{it} = 0)} = \exp(\beta^{PA}), \quad (6.4)$$

com $\exp(\beta^{PA})$ podendo ser interpretado como a razão da chance de uma criança da população ter uma doença respiratória quando a mãe é fumante comparativamente à chance de ter uma doença respiratória quando a mãe não é fumante.

Dessa maneira, Hardin e Hilbe (2013, p. 61) aponta para o cuidado que se deve ter acerca de qual o parâmetro de interesse.

Ambos os modelos estão relacionados com problemas de âmbito longitudinal, de modo que, se por um lado, o foco residisse sobre questões populacionais, os modelos PA seriam uma possibilidade estratégica, enquanto que, por outro lado, se as questões do pesquisador residissem mais sobre as características da heterogeneidade das UEs, os modelos SS seriam mais apropriados.

Do ponto de vista longitudinal, algumas UEs poderiam fornecer respostas (médias) acima ou, eventualmente, abaixo da média global, de modo que cada UE possa expressar um certo padrão em suas próprias respostas ao longo do tempo. Isso, de certa forma, abre espaço para estudos e questionamentos acerca da variabilidade (heterogeneidade) existente entre as UEs, denominada *efeitos aleatórios*.

Em estudos longitudinais, porém, se as questões fundamentais estivessem baseadas no padrão da resposta média populacional (p. ex., o caso de estudos epidemiológicos), então, a estratégia passaria pela implementação de um modelo do tipo PA.

De um modo geral, o objetivo do artigo de Zeger et al. (1988) consistia em explorar as possíveis conexões entre modelos PA e SS, tendo, em ambos os casos, o método de GEE de estimação. É importante enfatizar que esse artigo antecede em cinco anos o surgimento da classe de modelos GLMM. Daí a importância dos autores em estabelecer um mesmo método de estimação (GEE) adaptado para ambas as classes de modelos (PA e SS).

Os autores, nesse artigo, além de descrevem em detalhes as adaptações metodológicas para o processo de estimação de dois modelos (PA e SS), obtiveram uma relação matemática exata (resultado de um efeito de atenuação) entre esses mesmos modelos, para o caso de efeitos aleatórios com distribuição normal, com respostas discretas e contínuas.

Ao final, os autores testaram essa relação matemática (ou efeito de atenuação) com uma base de dados de doenças respiratórias em crianças (Harvard Study of Air Pollution and Health, Ware et al., 1984). Os resultados obtidos corroboraram as proposições dos autores.

Inicialmente, Zeger et al. (1988) descrevem o modelo SS considerando y_{it} e $\mathbf{x}_{it}$ como a variável resposta e o vetor de covariáveis fixas de ordem $p \times 1$, respectivamente, no tempo t e para a i -ésima UE, onde $t = 1, \dots, n_i$ e $i = 1, \dots, K$.

Além disso, considerou $\mathbf{z}_{it}$ um vetor $q \times 1$ de covariáveis (possivelmente um subconjunto de $\mathbf{x}_{it}$), associado com o vetor (também $q \times 1$) de coeficientes dos efeitos aleatórios, $\mathbf{b}_i$. Consideraram como as respostas médias condicionais da i -ésima UE

$$u_{it} = E(y_{it}|\mathbf{b}_i), \quad (6.5)$$

devendo satisfazer

$$h(u_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i, \quad (6.6)$$

com $Var(y_{it}|\mathbf{b}_i) = g(u_{it}) \cdot \phi$, sendo h e g as funções de ligação e de variância, respectivamente. Note, ainda, que a função de variância, g , é descrita como uma função dependente da média, $g(u_{it})$, característica incorporada da classe GLM.

A título de ilustração, Zeger et al. (1988) mencionam uma regressão logística de modelos mistos e efeitos aleatórios com distribuição normal, então descrita por Stiratelli et al. (1984), com a forma

$$\text{logit}(u_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i \quad e \quad Var(y_{it}|\mathbf{b}_i) = u_{it}(1 - u_{it}), \quad (6.7)$$

com $Var(y_{it}|\mathbf{b}_i) = u_{it}(1 - u_{it})$ e $\mathbf{b}_i$ sendo um vetor aleatório independente com distribuição $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$.

Além disso, Zeger et al. (1988) consideraram que em uma situação específica do tipo $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t)'$ com $\mathbf{b}_i = (b_{0i}, b_{1i})'$ e $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, a estrutura de um modelo logístico misto para descrever o logaritmo da razão de chances de uma resposta positiva para uma UE no instante t , seria dada por $(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t$, que é função linear do tempo t .

É importante notar que a variável t (tempo) estaria exercendo a função de variável independente. Nesse caso, a interpretação de seu coeficiente associado é descrita pelos dois itens a seguir:

- a quantidade $\exp(\beta_1 + b_{1i})$ é a razão de chances de uma resposta positiva em $t + 1$, relativamente a t , para a i -ésima UE;
- como $\mathbb{E}(b_i) = 0$, o parâmetro β_1 descreveria, em média, a forma como a probabilidade de resposta positiva da i -ésima UE é influenciada pelo tempo.

Em contrapartida, em um contexto de modelo PA, a média marginal $\mu_{it} = \mathbb{E}(y_{it})$ constitui o foco. De forma a diferenciar o modelo SS, os autores marcaram as funções h , g e o vetor $\boldsymbol{\beta}$ com uma estrela, isto é,

$$h^*(\mu_{it}) = \mathbf{x}_{it}' \boldsymbol{\beta}^*, \quad (6.8)$$

com $\text{Var}(y_{it} | \mathbf{b}_i) = g^*(\mu_{it}) \cdot \phi$.

Note, ainda, que para o modelo SS, os autores estão usando u_{it} (em $h(u_{it})$ e $g(u_{it})$), enquanto que, para o modelo PA, estão usando μ_{it} (em $h^*(\mu_{it})$ e $g^*(\mu_{it})$).

Nesse caso, $\boldsymbol{\beta}^*$ expressa a forma como a resposta média da população (e não de alguma UE) depende das covariáveis.

De forma semelhante ao caso anterior, isto é, se fosse admitido $\mathbf{x}_{it} = \mathbf{z}_{it} = (1, t)'$, $\mathbf{b}_i = (b_{0i}, b_{1i})'$ e $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1^*)'$, a quantidade β_1^* passaria a representar, na escala *logit*, a mudança na fração de respostas positivas por unidade de tempo (e não a mudança de alguma UE em particular).

Zeger et al. (1988) citam dois grandes aspectos capazes de distinguir os modelos PA e SS:

1. A primeira (e a principal) diz respeito às possibilidades de interpretação dos coeficientes do modelo de regressão, no sentido de refletirem mudanças na resposta média da população ou, de outro modo, de uma UE em particular, frente a mudanças em $\mathbf{x}$;
2. A natureza da dependência temporal assumida.

O primeiro aspecto remete, de certa forma, a questões de delimitação e, portanto, do reconhecimento das especificidades e peculiaridades de cada um dos modelos.

Com relação ao segundo aspecto, os autores chamam a atenção para três pontos fundamentais que, por estarem interconectados, ajudariam a compreender ainda mais essa distinção em uma perspectiva mais fundamental. São eles:

- Enquanto que modelos PA descreveriam a covariância entre medidas repetidas de uma mesma UE, os modelos SS explicariam essa fonte de covariância.
- Em modelos PA, a matriz de covariância deve ser positiva definida, enquanto que, nos modelos SS, a dependência temporal surge unicamente dos efeitos aleatórios (compartilhados em nível individual) $\mathbf{b}_i$, de modo que a matriz de covariância de um

modelo SS é totalmente determinada pelas escolhas da função de variância, $g(u_{it})$, e de F (distribuição de $\mathbf{b}_i$).

Zeger et al. (1988) chamam também a atenção para o fato de que o uso de um modelo SS é limitado pela informação disponível das UEs. Como nem sempre a quantidade de informação de cada UE é abundante, assumir que $\mathbf{b}_i$ segue uma distribuição de probabilidades funciona como uma espécie de “. . . a vehicle for borrowing strength across subjects to estimate β , the typical SS parameter.” (p. 1051)

Uma vantagem dos modelos PA é que uma resposta média da população, para um dado vetor de covariáveis $\mathbf{x}$, é perfeitamente passível de ser obtida sem a suposição de heterogeneidade das UEs. Nesse sentido, os parâmetros de um modelo PA são mais próximos dos dados, mas acaba dependendo do grau de heterogeneidade na população (F).

Para notarmos esse fato, retomamos (6.5) e (6.6). Como $h(u_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i$, então, por meio de sua função inversa, podemos obter $u_{it} = h^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i)$, de modo que a média marginal μ_{it} , a partir da média condicional u_{it} , passa a ser dada por

$$\mu_{it} = \mathbb{E}(y_{it}) = \mathbb{E}[\mathbb{E}(y_{it}|\mathbf{b}_i)] = \int h^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i) dF(\mathbf{b}_i), \quad (6.9)$$

sendo possível notar a dependência de μ_{it} tanto de $\mathbf{x}_{it}$ quanto de F .

Em seguida, Zeger et al. (1988) abordam o problema fundamental a ser analisado nesta Seção 6.2.1, a saber, a questão da *atenuação dos efeitos das covariáveis nos parâmetros*.

Tendo em mente um modelo geral tal como em (6.9), os autores notaram que:

- A média marginal μ_{it} depende tanto de $\mathbf{x}_{it}$ quanto de F ;
- Se a função de ligação h é não linear e assumirmos $h(u_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i$, então, não é necessariamente verdadeiro que $h(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta}$, isto é, que a partir de (6.6) não se chega necessariamente em (6.8);
- Se, por ventura, não existir heterogeneidade, isto é, se $\mathbf{b}_i = \mathbf{0}$, para todo i , então, os modelos PA e SS serão um mesmo modelo, isto é, $h(u_{it}) = h(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta}$;
- Ainda nessa mesma linha de raciocínio, se h for a função de ligação identidade, então, $h(\mu_{it}) = \mu_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}$;
- Em geral, a função de ligação que transforma a média condicional u_{it} em uma função que seja linear das covariáveis, $\mathbf{x}_{it}$, não necessariamente garante o mesmo tipo de transformação para a média marginal μ_{it} .

Para ilustrar esse último item, os autores propuseram um modelo logístico para a média condicional

$$\text{logit}(u_{it}) = \eta_{it} + b_i, \quad b_i \sim N(\mathbf{0}, \mathbf{D}), \quad (6.10)$$

a partir do qual se obtém a média marginal μ_{it} , como função de η_{it} , para vários valores de $\mathbf{D}$. O comportamento gráfico da média marginal encontra-se na Figura 6.1.

Figura 6.1 – Média marginal μ_{it} vs. η_{it} , com Y satisfazendo $\text{logit}[E(Y|\mathbf{b})] = \eta + \mathbf{b}$

Fonte: Zeger, Liang and Albert (1988, p. 1052).

Com base na Figura 1, é possível perceber a dependência da média marginal μ_{it} com relação à variância dos efeitos aleatórios.

De acordo, ainda, com Zeger et al. (1988), esse fenômeno é conhecido como *atenuação do efeito das covariáveis*, visualmente perceptível pela diminuição da inclinação das curvas do gráfico da Figura 6.1. Esse fenômeno também é conhecido e documentado na área de regressão com erro nas variáveis, com a citação de Stefanski (1985), por parte dos autores.

Zeger et al. (1988) obtiveram, ainda, alguns resultados importantes para as expressões das médias marginais μ_{it} , em função de algumas funções de ligação conhecidas. Esses resultados são restritos para o caso da distribuição F (efeitos aleatórios) com distribuição normal, média $\mathbf{0}$ e matriz de covariância $\mathbf{D}$ e encontram-se na Tabela 6.1.

Esses resultados da Tabela 6.1, obtidos por Zeger et al. (1988), dizem respeito à primeira tentativa de expressar uma relação matemática exata entre os parâmetros dos modelos PA e SS ou, mais especificamente, entre β^{SS} e β^{PA} . Esses resultados são de grande importância nesta tese e serão utilizados e analisados em detalhes no Capítulo 7.

Tabela 6.1 – Relação entre média condicional, função de ligação e média marginal para classe de respostas GLM

Função de ligação	Média marginal
Identidade [$h(u) = u$]	$\mu_{it} = \mathbb{E}(y_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta}$
Log [$h(u) = \log(u)$]	$\mu_{it} = \exp\left(\mathbf{x}'_{it}\boldsymbol{\beta} + \frac{\mathbf{z}'_{it}\mathbf{D}\mathbf{z}_{it}}{2}\right)$ Obs.: os efeitos aleatórios produzem uma compensação (<i>offset</i>) na média marginal.
Probit [$h(u) = \Phi^{-1}(u)$]	$\mu_{it} = \Phi(a_p(\mathbf{D}) \cdot \mathbf{x}'_{it}\boldsymbol{\beta}),$ com $a_p(\mathbf{D}) = \mathbf{D}\mathbf{z}_{it}\mathbf{z}'_{it} + \mathbf{I} ^{-q/2}$ e q sendo a dimensão de $\mathbf{b}_i$. Obs.: a expressão $a_p(\mathbf{D})$ é uma forma generalizada do parâmetro s em Gilmour et al. (1985).
Logit (resultado de Johnson e Kotz (1970, p. 6))	$\text{logit}(\mu_{it}) \approx a_l(\mathbf{D}) \cdot \mathbf{x}'_{it}\boldsymbol{\beta},$ com $a_l(\mathbf{D}) = c^2\mathbf{D}\mathbf{z}_{it}\mathbf{z}'_{it} + \mathbf{I} ^{-q/2},$ com $c = 16\sqrt{3}/(15\pi).$

Fonte: Zeger, Liang and Albert (1988, p. 1054).

Uma importante menção deve ser feita: não foi encontrada, a partir da média condicional, uma forma fechada para a expressão da média marginal para o caso da função de ligação *logit*. No entanto, os autores utilizaram um resultado de Johnson e Kotz (1970, p. 6) ao aproximar a função logística pela função de distribuição acumulada normal.

De acordo, ainda, com os autores, as variáveis respostas podem ser acomodadas segundo um modelo GLM misto, isto é, variável resposta com distribuição pertencente à família exponencial, permitindo estrutura com efeitos fixos e aleatórios.

Como aplicação, Zeger et al. (1988) citam um estudo de Ware et al. (1984), da Harvard Study of Air Pollution and Health, no qual 537 crianças foram examinadas anualmente e a relação entre infecção respiratória da criança e a condição de fumante da mãe foi estudada. A variável resposta compreende o relato da própria mãe no que diz respeito ao filho(a) ter tido infecção respiratória no ano anterior ao exame (0 ou 1). O *status* da mãe (MS) também é variável binária e a variável AGE foi transformada de modo que $AGE = 0$ diz respeito a crianças a partir dos 9 anos de idade.

O modelo PA proposto para a probabilidade marginal da infecção respiratória, μ_{it} , é dado por

$$\text{logit}(\mu_{it}) = \beta_0^* + MS \cdot \beta_1^* + AGE \cdot \beta_2^* + (AGE \cdot MS)\beta_3^*. \quad (6.11)$$

Os resultados do modelo PA seguem na Tabela 6.2.

Tabela 6.2 – Resultados do Modelo PA – Harvard Study of Air Pollution and Health
Coefficients and robust Z-statistics ($\hat{\beta}^/s.e.\hat{\beta}^*$) for the population-averaged model of equation (4.1) for three different assumptions about the correlation among repeated observations for a subject*

Coefficient	Uncorrelated		Exchangeable ^a		Stationary ^b	
	$\hat{\beta}^*$	Z	$\hat{\beta}^*$	Z	$\hat{\beta}^*$	Z
Intercept	-1.90	-16.0	-1.90	-15.8	-1.90	-15.8
Mother's smoking (MS = 0 if no)	.313	1.7	.303	1.6	.298	1.6
AGE	-.140	-2.4	-.137	-2.4	-.139	-2.4
MS · AGE	.0699	.79	.0657	.75	.0704	.80

^a $\hat{\alpha} = .346$.

^b $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3) = (.40, .31, .30)$.

Fonte: Zeger, Liang and Albert (1988, p. 1057).

Os autores utilizaram três diferentes suposições de correlação para demonstrar que tanto $\hat{\beta}^*$ quanto as estatísticas Z não dependem tanto da escolha da matriz de correlação. São elas:

- *Uncorrelated*: $\mathbf{R} = \mathbf{I}$;
- *Exchangeable*: $R_{jk} = \alpha$, para $j \neq k$;
- *Stationary Correlation*: $R_{jk} = \alpha(|j - k|)$.

Assim, para uma estrutura de correlação *exchangeable*, $\hat{\alpha} = 0,346$. Já para uma estrutura de correlação *stationary*, os lags 1, 2 e 3 apresentaram correlações estimadas em 0,40, 0,31 e 0,31, respectivamente.

Com relação às estimativas dos parâmetros, entre as crianças de 9 anos ($AGE = 0$) com mães não fumantes, a taxa anual de infecção respiratória é de aproximadamente 15% ($e^{-1,9}$), nos três modelos.

A taxa de doença de crianças com mães fumantes é 1,35 maior quando comparada com a taxa de crianças com mães não fumantes. Uma vez que as estimativas dos três modelos PA ficaram bem próximas, basta fazer (em qualquer um dos três modelos) $e^{\hat{\beta}_0^* + \hat{\beta}_1^*} / e^{\hat{\beta}_0^*}$.

Como as estatísticas Z, para β_1^* , ficaram todas entre 1,6 e 1,7, os autores sugerem evidências apenas moderadamente favoráveis para relação entre a condição de fumante da mãe e a possibilidade de infecção respiratória da criança.

Com relação ao modelo SS (para o mesmo conjunto de dados), os autores sugeriram o modelo

$$\text{logit}(u_{it}) = \beta_0 + MS \cdot \beta_1 + AGE \cdot \beta_2 + (AGE \cdot MS)\beta_3 + b_{0i}, \quad (6.12)$$

com $b_{0i} \sim N(0, D)$, de forma que apenas o intercepto varia entre as crianças.

Para esse modelo, Zeger et al. (1988) sugerem interpretar os coeficientes do modelo como o logaritmo da razão de chances para uma criança. Assim, por exemplo, o coeficiente β_1 diz respeito a quanto mudaria o risco de infecção de uma criança se a condição (*status* de fumante) da mãe fosse alterada.

Os resultados do modelo S,S estimado pelos autores encontram-se na Tabela 6.3, para diferentes valores da variância dos efeitos aleatórios.

Tabela 6.3 – Resultados do Modelo SS – *Harvard Study of Air Pollution and Health*
Coefficients and robust Z-statistics from subject-specific model with a Gaussian random intercept with variance D

Coefficient	D							
	0		.5		1		4.0	
	$\hat{\beta}$	Z	$\hat{\beta}$	Z	$\hat{\beta}$	Z	$\hat{\beta}$	Z
Intercept (β_0)	-1.91	-16	-2.1	-16	-2.2	-17	-2.9	-16
$MS(\beta_1)$	.313	1.6	.344	1.7	.368	1.7	.488	1.6
$AGE(\beta_2)$	-.14	-2.4	-.15	-2.5	-.16	-2.5	-.22	-2.5
$MS \cdot AGE(\beta_3)$	.070	.79	.074	.78	.079	.77	.11	.78
$a_l(D)$	1.00		1.08		1.16		1.54	
$\hat{\beta}_1(0) \cdot a_l(D)$	.313		.338		.363		.488	
$\log \left[\frac{\hat{\beta}_1(0) \cdot a_l(D)}{\hat{\beta}_1(D)} \right]$	0.0		-.02		-.01		-.01	

Fonte: Zeger, Liang and Albert (1988, p. 1058).

Inicialmente devemos notar que, quando $D = 0$, algumas estimativas do modelo SS são ligeiramente superiores às do modelo PA (β_0 e β_3). No entanto, para vários valores de D , todas as estimativas do modelo SS são (em módulo) maiores que as estimativas do modelo PA. Conforme observação de Zeger et al. (1988), a variância dos efeitos aleatórios, D , “encolhe” os parâmetros dos efeitos fixos em um contexto de regressão logística, condizendo, portanto, o contexto com o efeito da atenuação das covariáveis.

Segundo ainda os autores, o grau do encolhimento depende do estudo do planejamento dos $\mathbf{z}_{it}$'s, mas que, de acordo com os autores, pode ser aproximado por $a_l(D)$, resultado obtido por Johnson e Kotz (1970) cuja expressão encontra-se na Tabela 6.1. Quando D cresce, as estimativas (aproximadas) para o grau de encolhimento são crescentes (1,00, 1,08, 1,16 e 1,54), o que produz evidências acerca do fenômeno da atenuação do efeito das covariáveis.

Ao avaliarem esse efeito sobre o parâmetro (fixo) β_1 , os autores estimaram as quantidades $\hat{\beta}_1(0) \cdot a_l(D)$, para D variando de 0 a 4 (0, 0,5, 1,0 e 4,0), sendo $\hat{\beta}_1(0)$ a estimativa do parâmetro β_1 quando $D = 0$. Esses valores foram estimados em 0,313, 0,338, 0,363 e 0,488, respectivamente.

Quando comparados contra as respectivas estimativas de $\hat{\beta}_1(\mathbf{D})$, via $\log \left[\frac{\hat{\beta}_1(0) \cdot a_l(\mathbf{D})}{\hat{\beta}_1(\mathbf{D})} \right]$, essas diferenças tendem a zero.

Outra característica apontada pelos autores diz respeito ao fato de as estatísticas Z praticamente não mudarem de valor, ainda que para valores crescentes de $\mathbf{D}$. Mas, conforme descrito anteriormente, as estimativas de β mudaram substancialmente, em termos absolutos.

Em consonância com o modelo PA, os valores Z sugerem evidências apenas moderadamente favoráveis para relação entre a condição de fumante da mãe e a possibilidade de infecção respiratória da criança, tendo a variável idade também apresentado significância (para todos os valores de $\mathbf{D}$).

De um modo geral, Zeger et al. (1988) apontaram que, por meio do modelo SS, o risco de infecção de uma criança poderia diminuir de 35% ($e^{-1,91+0,313}/e^{-1,91}$), quando $\mathbf{D} = \mathbf{0}$, percentual semelhante ao modelo PA) até 63% ($e^{-2,9+0,488}/e^{-2,9}$), à medida que a mãe parasse de fumar (mudança de *status*), levando, ainda, em consideração uma variabilidade ($\mathbf{D}$) entre 0 e 4 para os efeitos aleatórios.

6.2.2 Os Modelos PA e SS – Alguns Resultados Gerais para Regressão Logística

Nesta seção, analisaremos algumas inovações (bem como os resultados oriundos dessas inovações) proporcionadas pelo artigo de Neuhaus, Kalbfleisch e Hauck (1991) com relação aos modelos PA e SS, ainda em um contexto de regressão logística.

No caso, os autores propõem o modelo apenas com uma variável explicativa, mas asseguram que os resultados do artigo podem ser estendidos para o caso de múltiplas variáveis (explicativas).

Fundamentalmente, os autores obtiveram aproximações para β , em um contexto de regressão logística, por meio da minimização da divergência de Kullback-Leibler. Em seguida, apresentou um teorema, o qual permite realizar comparações entre os parâmetros β dos modelos PA e SS a partir dessa nova metodologia. A classe de modelos SS considerada pelos autores é *cluster-specific*.

Neuhaus et al. (1991) constataam que seus resultados são bastante similares aos de Zeger et al. (1988), ainda que obtidos por um método distinto. A título de ilustração, os resultados de Neuhaus et al. (1991) permitem utilizar qualquer distribuição para os efeitos aleatórios.

Neuhaus et al. (1991) partem de um modelo logístico de efeitos mistos (com função de ligação *logit*), com intercepto aleatório α_i variando pelas UEs com função densidade $f(\alpha)$. O modelo condicional assume, portanto, a forma

$$\text{logit } \mathbb{P}(Y_{ij} = 1 | \alpha_i, X_{ij}) = \alpha_i + \beta X_{ij}. \quad (6.13)$$

A interpretação do coeficiente β , segundo Neuhaus et al. (1991), é descrita como a mudança na probabilidade (com função de ligação *logit* e levando em consideração a covariável X) de resposta, para UEs sujeitas ao grupo de risco descrito por α_i . De certa maneira, essa interpretação é equivalente à de Zeger et al. (1988), uma vez que enfatiza (em função de α_i) a probabilidade de resposta da UE (e não a da população).

Contudo, a abordagem GEE que caracteriza a distribuição marginal de Y_{ij} (classe de modelos PA) é descrita por

$$\text{logit } \mathbb{P}(Y_{ij} = 1 | X_{ij}) = \alpha^* + \beta^* X_{ij}, \quad (6.14)$$

com alguma estrutura de covariância para Y .

A interpretação também segue na mesma linha de raciocínio de Zeger et al. (1988), isto é, β^* mede a mudança na escala *logit* da proporção de respostas positivas ($Y = 1$), para uma unidade de acréscimo em X .

O modelo logístico de efeitos mistos (6.13) é capaz de especificar um modelo marginal para Y_{ij} ao obtermos a integral

$$\mathbb{P}(Y_{ij} = 1 | X_{ij}) = \int (1 + e^{-\alpha - \beta X_{ij}})^{-1} f_{\theta}(\alpha) d\alpha, \quad (6.15)$$

ainda que sua forma não seja, precisamente, a de um modelo logístico.

Inicialmente, Neuhaus et al. (1991) consideram que o efeito do aumento em uma unidade em X (na escala do log das chances) é dado por

$$\beta_{PA}(X) = \log \frac{\mathbb{P}(Y = 1 | X + 1) / \mathbb{P}(Y = 0 | X + 1)}{\mathbb{P}(Y = 1 | X) / \mathbb{P}(Y = 0 | X)}, \quad (6.16)$$

sendo que, em um contexto de modelo PA, o parâmetro $\beta_{PA}(X) = \beta^*$ e independente de X . Essa relação de independência pode ser notada na própria interpretação de (6.14).

Para a classe de modelos SS, $\beta_{PA}(X)$ depende de X , fato que pode ser notado na interpretação de (6.13), e, a partir de (6.15) e (6.16), temos que

$$\beta_{PA}(X) = \log \left(\frac{\mathbb{E} \left\{ (1 + e^{-\alpha - \beta(x+1)})^{-1} \right\} \mathbb{E} \left\{ (1 + e^{\alpha + \beta x})^{-1} \right\}}{\mathbb{E} \left\{ (1 + e^{\alpha + \beta(x+1)})^{-1} \right\} \mathbb{E} \left\{ (1 + e^{-\alpha - \beta x})^{-1} \right\}} \right), \quad (6.17)$$

sendo que as esperanças são avaliadas com relação a $f_{\theta}(\alpha)$.

Como o processo de integração em (6.17) é de difícil obtenção, Neuhaus et al. (1991) aproximaram o lado direito de (6.17) (após efetuarem a expansão de Taylor em torno de $\beta = 0$ e realizarem algumas simplificações) pela expressão

$$\beta_{PA}(X) \approx \beta \left\{ 1 - \frac{Var(p)}{\mathbb{E}(p)\mathbb{E}(q)} \right\} = \beta[1 - \rho(0)], \quad (6.18)$$

com $\text{logit}(p) = \alpha$, $q = 1 - p$ e $\rho(0) = \text{Corr}(Y_{ij}, Y_{ij'} | \beta = 0)$, para o modelo logístico de efeitos mistos. O termo $\rho(0)$ consiste na correlação intraclasse, quando a covariável possui efeito nulo.

Os autores chamam a atenção para o fato de a equação (6.18) também poder ser obtida via Teoria da Informação, isto é, pela divergência de Kullback-Leibler. Nesse caso, Neuhaus et al. (1991) chegaram na expressão

$$\alpha^* + \beta^* X \approx \log \left[\frac{\mathbb{E}(p)}{\mathbb{E}(q)} \right] + \beta X \left[1 - \frac{Var(p)}{\mathbb{E}(p)\mathbb{E}(q)} \right], \quad (6.19)$$

a partir do qual estimativas consistentes dos parâmetros do modelo (6.14) convergem para aqueles que minimizam a divergência Kullback-Leibler, isto é, para α^* e β^* .

Assim, Neuhaus et al. (1991) ressaltam que o modelo misto induzido via (6.15) é aproximadamente logístico com coeficiente β atenuado dado pela expressão (6.18).

A partir dos resultados (6.18) e (6.19), Neuhaus et al. (1991) fizeram algumas considerações de grande importância:

- Uma vez que a correlação intraclasse $0 \leq \rho(0) \leq 1$, então, ao menos para um β pequeno, $|\beta_{PA}(X)| < |\beta|$, de modo que o efeito medido pelo modelo PA é menor que o efeito medido pelo modelo SS;
- O resultado (6.18) sugere que, quanto maior $Var(p)$ (ou, equivalentemente, $Var(\alpha)$), menor o valor de β , e, portanto, maior a diferença entre β_{PA} e β ;
- Se os efeitos aleatórios são nulos, então, $Var(p) = 0$ e $\beta = \beta_{PA}(X)$, e o modelo SS (6.13) também especifica o modelo PA (6.14).

Uma menção sobre o primeiro item: esse item passou a receber, por parte dos autores, um tratamento mais formal, tendo sido apresentado, mais adiante, sob a forma de teorema.

Os autores notaram, ainda, que Zeger et al. (1988) também forneceram, mesmo que apenas para o caso particular de efeitos aleatórios com distribuição normal, uma medida de atenuação qualitativamente similar à deles, isto é, que os parâmetros estimados de um modelo PA tendem a ficar mais próximos de zero quando comparados com os parâmetros de um modelo

SS. Mas, a abordagem trazida por Neuhaus et al. (1991) teria, na própria expressão (6.14), um apelo mais intuitivo.

A título de ilustração, Neuhaus et al. (1991) apresentam esses resultados por meio de uma base de dados referente a um estudo da Universidade da Califórnia, São Francisco, sobre fluidos colhidos do(s) seio(s) de 490 mulheres brancas, em fase de pré-menopausa e que não tiveram doenças nos seios. Os autores citam Ernster et al. (1987) para mais detalhes.

Se os fluidos foram obtidos de ambos os seios, então, $Y = 1$. Caso contrário, $Y = 0$. As covariáveis consideradas foram:

- Evidências (por meio de exames) de displasia, $X_1 = 1$; caso contrário, $X_1 = 0$;
- Idade em anos (X_2);
- Idade na menarca (X_3);
- Variável indicadora. Se já deu à luz, $X_4 = 1$; caso contrário, $X_4 = 0$.

Foram ajustados dois modelos: i) um modelo SS, representado em (6.13), e ii) um modelo PA, representado em (6.14), com estrutura de correlação *exchangeable* e estimadores robustos. Os resultados encontram-se na Tabela 6.4.

De um modo geral, todas as estimativas do modelo PA são, em módulo, menores, isto é, mais próximas de zero quando comparadas com as estimativas do modelo SS, resultado que corrobora o prognóstico dos autores descrito na expressão (6.18).

Tabela 6.4 – Estimativas dos modelos PA e SS para dados de doenças nos seios

Point estimates (standard errors) of regression coefficients for cluster-specific and population-averaged approaches applied to data on the availability of breast fluid.

Variable	Mixed effects, β_{CS}	Liang-Zeger, β_{PA}	Ratio, β_{PA}/β_{CS}
Intercept	-1.951 (2.49)	-0.620 (0.908)	
Dysplasia	0.609 (0.462)	0.208 (0.155)	0.351
Age	0.146 (0.044)	0.050 (0.015)	0.343
Age at menarche	-0.414 (0.172)	-0.147 (0.059)	0.355
Full term birth	1.249 (0.610)	0.435 (0.216)	0.348
std. dev. (α_i)	4.218 (0.511)		

Fonte: Neuhaus, et al. (1991, p. 32).

No modelo PA, mulheres que deram à luz exibiram *logit* de prevalência de fluidos 0,435 unidade maior quando comparadas com mulheres nulíparas.

Em contrapartida, o modelo SS sugere que, entre mulheres de um mesmo grupo de risco α_i , o *logit* da prevalência dos fluidos é 1,25 unidade maior entre aquelas que já deram à luz.

A Tabela 6.4 fornece também a razão entre β_{PA} e β_{SS} , de modo a se verificar (mais uma vez) a aproximação de $\left\{1 - \frac{Var(p)}{E(p)E(q)}\right\}$, na expressão (6.18), a qual foi estimada em 0,35 pelos autores por meio da distribuição de α . Ao comparar esse valor com os valores dispostos na última coluna da Tabela 6.4, a aproximação é evidente. Segundo, ainda, os próprios autores, a aproximação pelo método de Zeger et al. (1988) também teria sido expressiva, com estimativa igual a 0,37.

Por fim, Neuhaus et al. (1991) ainda chamam atenção para o fato de que os valores padronizados $(\hat{\beta}/se(\hat{\beta}))$ em ambos os modelos são bem próximos, de modo que os testes do tipo Wald acabam, também, por produzir níveis de significância próximos.

É interessante destacar alguns aspectos de ordem temporal. Raudenbush (2009, p. 168) destaca que, somente em 2000, Heagerty e Zeger (2000) conseguiram desvincular completamente o método de estimação da escolha de modelo (SS ou PA). Ao apresentarem um modelo (multinível) para a média marginal via estimação de máxima verossimilhança, Heagerty e Zeger (2000) permitiram que os modelos PA não mais estivessem restritos a um único método de estimação (GEE). Apesar da relevância do artigo de Heagerty e Zeger (2000), o foco não consistia em estudar as relações entre os modelos PA e SS.

Dessa maneira, daremos sequência com as análises das relações entre esses modelos (PA e SS), destacando modelos longitudinais com respostas de Poisson, no intuito de explorar as possibilidades de existência de relações exatas.

6.3 Os modelos PA e SS – Regressão de Poisson (parte I)

Antes de dar prosseguimento à análise do artigo de Young, Preisser, Qaqish e Wolfson (2007), talvez seja interessante abordar a estrutura matemática de um modelo SS padrão, isto é, de um modelo longitudinal com resposta pertencente à família exponencial e a efeitos aleatórios.

Uma vez que procuramos modelos com o maior grau de opções ao pesquisador, em termos de distribuição da variável resposta, a classe de modelos GLMM é de interesse, uma vez que ela também é extensão da classe de modelos GLM, a qual acomoda qualquer distribuição de probabilidades pertencente à família exponencial.

Como será analisado mais adiante, parte considerável do problema de se encontrar equivalência entre os modelos SS e PA reside no estabelecimento adequado de funções de

ligação. Uma vez que se busca uma maior flexibilidade para a variável resposta, tanto maior tende a ser a complexidade das funções de ligação (funções de ligação não lineares).

A título de ilustração, os modelos de regressão com resposta binária (regressão logística) necessitam de uma função de ligação que, de fato, estabeleça a relação entre os valores médios a serem estimados (baseados nas respostas Y_{ij} , zero e um) e o preditor linear representando uma combinação entre efeitos fixo e aleatório $X'_{ij}\beta + Z'_{ij}b_i$. Logo, uma vez que $0 < \mu_{ij} < 1$, uma função de ligação deve satisfazer tal condição, e funções de ligação tais como *logit* e *probit* atendem a essa exigência.

Uma outra ilustração também conhecida diz respeito ao modelo de regressão com resposta de Poisson e efeitos aleatórios. Esse tipo de modelo também necessita de uma função de ligação que, de fato, estabeleça a igualdade entre os valores médios a serem estimados, nesse caso, baseados em valores positivos (a variável resposta, Y_{ij} , refere-se, nesse caso, a dados de contagem) e a soma dos efeitos fixos e aleatórios $X'_{ij}\beta + Z'_{ij}b_i$.

Ainda sobre a regressão de Poisson, o fato da variável resposta expressar contagens, a média da variável resposta deverá ser sempre positiva, isto é, $\mu_{ij} > 0$. Além disso, a possibilidade de existência de efeitos multiplicativos faz com que uma função de ligação *log*, sob a forma $\eta_{ij} = \log(\mu_{ij})$, seja adequada para esse tipo de situação.

Esses dois exemplos ilustrativos demonstram como a função de ligação *identidade* (tão utilizada em modelos com variável resposta normal) não é tão profícua quando a distribuição da variável resposta não possui distribuição normal. De forma a abordar a estrutura desse problema, consideremos um modelo condicional cuja função densidade de probabilidade seja descrita sob a forma da família exponencial

$$f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}) = \exp\{[y_i\gamma_i - b(\gamma_i)]/\phi^2 - c(y_i, \phi)\}, \quad (6.20)$$

com $y_i|\mathbf{u}$ independentes entre si e os efeitos aleatórios com uma determinada distribuição de probabilidades, isto é, $\mathbf{u} \sim f_U(\mathbf{u})$.

Para distribuições pertencentes à família exponencial, é verdadeiro que $\mu_i = \frac{\partial b(\gamma_i)}{\partial \gamma_i}$, de modo que, da perspectiva do modelo em questão, $\mathbb{E}(y_i|\mathbf{u}) = \mu_i$. Conforme recém-ilustrado (nos modelos de regressão logística e de Poisson), a função de ligação deve estabelecer uma relação de igualdade entre a resposta (lado esquerdo do sinal de igual) com alguma combinação entre efeitos fixos e aleatórios (lado direito do sinal de igual). Essa função de ligação é, na prática, obtida por meio de alguma transformação da média. Em outras palavras, a função de

ligação assumirá o papel de uma função g da média, $g(\mu_i)$, podendo assumir várias formas que sejam de interesse. Assim, a função de ligação, em um modelo SS, estabelece que

$$g(\mu_i) = X'_{ij}\beta + Z'_{ij}b_i. \quad (6.21)$$

Assim, por exemplo, se considerarmos a função de ligação mais simples (função de ligação *identidade*), temos que $g(\mu_i) = \mu_i$, fazendo com que $\mu_i = X'_{ij}\beta + Z'_{ij}b_i$.

Além disso, um mesmo modelo de regressão pode acomodar mais de uma função de ligação. A própria Tabela 6.1 ilustra esse fato. A regressão logística com efeitos aleatórios contava com quatro possíveis funções de ligação: *identidade*, *log*, *probit* e *logit*.

Para simplificar a notação a seguir, vamos considerar $g(\mu_i) = x'_i\beta + z'_ib_i$. A média de y_i poderá ser obtida pelo processo iterativo de esperanças (esperança da esperança), ou seja,

$$\begin{aligned} \mathbb{E}(y_i) &= \mathbb{E}[\mathbb{E}(y_i|\mathbf{u})] \\ &= \mathbb{E}[\mu_i]. \end{aligned} \quad (6.22)$$

Em um modelo com função de ligação, porém, a obtenção de μ_i depende do cálculo de inversão da função de ligação $g(\mu_i)$, de modo que $\mu_i = g^{-1}(x'_i\beta + z'_ib_i)$, e, dessa forma, teremos

$$\begin{aligned} \mathbb{E}(y_i) &= \mathbb{E}[\mu_i] \\ &= \mathbb{E}[g^{-1}(x'_i\beta + z'_ib_i)]. \end{aligned} \quad (6.23)$$

Segundo McCulloch e Searle (2001, p. 223), a equação (6.23) não é de fácil simplificação, principalmente quando a função de ligação g for não linear.

Vamos assumir uma função de ligação log, ou seja, $g(\mu) = \log \mu$. Uma vez que, por meio de sua inversa, $\mu = g^{-1}(\log \mu) = \exp\{\mu\}$, temos que

$$\begin{aligned} \mathbb{E}(y_i) &= \mathbb{E}[g^{-1}(x'_i\beta + z'_ib_i)] \\ &= \mathbb{E}[\exp\{x'_i\beta + z'_ib_i\}] \\ &= \exp\{x'_i\beta\}\mathbb{E}[\exp\{z'_ib_i\}] \\ &= \exp\{x'_i\beta\}M_U(z'_i), \end{aligned} \quad (6.24)$$

onde, por definição, $M_U(z'_i) = \mathbb{E}[\exp\{z'_ib_i\}]$ é a f.g.m. (função geradora de momentos) de $\mathbf{u}$, avaliada em z'_i . Basta lembrar para isso que, de um modo geral, $M_X(t) = \mathbb{E}[e^{tX}]$.

Além disso, Agresti (2015, p. 309) ressalta que um modelo GLMM com resposta de Poisson e, especificamente, com função de ligação *identidade* também é possível, mas possui um problema estrutural. Com uma função de ligação *identidade*, existe o risco de que o preditor linear possa, eventualmente, vir a assumir valores negativos.

No entanto, ao designar uma distribuição de probabilidades específica para os efeitos aleatórios b_i , $\mathbb{E}(y_i)$ acaba por incorporar as suposições implícitas.

A título de ilustração, se uma variável aleatória $X \sim N(\mu, \sigma^2)$, então, a sua f.g.m. assumirá a forma

$$M_X(t) = \exp\{\mu t + \sigma^2 t^2 / 2\}. \quad (6.25)$$

Se, for o caso de $X \sim N(0, \sigma^2)$, então, a f.g.m. assumirá a forma

$$M_X(t) = \exp\{\sigma^2 t^2 / 2\}. \quad (6.26)$$

Uma vez que, para os efeitos aleatórios, é frequentemente assumido uma distribuição normal com média zero, ou seja, $u_i \sim N(0, \sigma_u^2)$, então, a f.g.m. de $\mathbf{u}$, avaliada em z'_i assumirá a forma

$$M_U(z'_i) = \exp\{z'_i \sigma_u^2 z_i / 2\}. \quad (6.27)$$

Em contrapartida, também é comum fazer a suposição de que cada linha da matriz Z tenha uma simples entrada igual a 1, com todo o resto igual a zero (McCullogh & Searle, 2001, p. 223). Logo, a f.g.m. assumirá a forma

$$M_U(z'_i) = \exp\{\sigma_u^2 / 2\}. \quad (6.28)$$

Logo, temos que, se a distribuição normal designada para os efeitos aleatórios condiz com as suposições assumidas em (6.27), então, a expressão (6.24) assumirá a forma

$$\mathbb{E}(y_i) = \exp\{x'_i \beta\} \exp\{z'_i \sigma_u^2 z_i / 2\}. \quad (6.29)$$

Contudo, se a distribuição normal designada para os efeitos aleatórios condiz com as suposições assumidas em (6.28), então, a expressão (6.24) assumirá a forma

$$\mathbb{E}(y_i) = \exp\{x'_i \beta\} \exp\{\sigma_u^2 / 2\}. \quad (6.30)$$

Para o caso (6.29), nenhuma condição foi imposta à matriz dos efeitos aleatórios, enquanto que, no caso de (6.30), houve uma imposição (todas as linhas da matriz Z possuem o valor 1 na entrada e zero nos demais elementos da linha), condizente com a escolha, por parte do pesquisador, de trabalhar com interceptos aleatórios. Logo, decorre de (6.30) que

$$\log \mathbb{E}(y_i) = x'_i \beta + \sigma_u^2 / 2. \quad (6.31)$$

Young et al. (2007) também citam, no início do artigo, McCullogh e Searle (2001) como referência para a busca de equivalência entre os parâmetros SS e PA em modelos *log-lineares* com distribuição de Poisson e efeitos aleatórios normalmente distribuídos.

Nessa linha de raciocínio, Young et al. (2007) procuraram explorar essas relações em dois tipos de modelos:

1. Pós-teste;
2. Pré-teste/Pós-teste.

Mas Young et al. (2007), porém, chamam atenção para algumas observações importantes sobre os modelos PA e SS. Inicialmente, Young et al. (2007) começam definindo e

especificando tanto o modelo SS quanto o PA, sendo que, neste último, apontam dois problemas técnicos quando da utilização do método GEE para estimar modelos SS:

1. O uso do estimador robusto;
2. A estrutura de correlação.

Com relação ao primeiro item, os autores citaram dois estudos: um de Murray, Hannan, Wolfinger, Baker and Dwyer (1998) e o outro de Kauermann e Carroll (2001). O primeiro relata que, para uma inferência válida, o estimador robusto de variância exige um grande número de UEs (mais de 50). Do contrário, corre-se o risco de se inflacionar a probabilidade de erro tipo I.

Em contrapartida, de acordo com Young et al. (2007), os autores Kauermann e Carroll (2001) teriam sugerido que, quando o tamanho das UEs é maior, relativamente ao número de UEs, as estimativas robustas da variância seriam menos eficientes do que as estimativas baseadas em modelo (*model-based*), isto é, quando tanto o modelo para a média quanto a matriz de correlação estão corretamente especificados. De acordo com os autores desse artigo, quando um modelo de quase-verossimilhança (proposto) está correto, a estimativa *sandwich* é frequentemente mais variável do que a estimativa de variância usual.

Sobre a relação entre os parâmetros SS e PA, Young et al. (2007) lembram da dificuldade de se estabelecer esse tipo de correspondência, mas, no caso em que a variável resposta admite distribuição de Poisson e os efeitos aleatórios são normalmente distribuídos, as dificuldades podem ser superadas.

Young et al. (2007) partem de um modelo GLMM com função de ligação *log*, dado por

$$\log\{\mathbb{E}(Y_{ij}|\mathbf{U}_i)\} = x'_{ij}\beta^{SS} + z'_{ij}\mathbf{U}_i, \quad (6.32)$$

com x_{ij} representando vetor coluna de covariáveis, β^{SS} o vetor paramétrico SS, z_{ij} um vetor coluna de planejamento (tipicamente um subconjunto de x_{ij}) sobre os quais os efeitos aleatórios são atribuídos. No caso, os autores chamam atenção para o fato da especificação de z_{ij} permitir que as observações Y_{ij} 's compartilhem os mesmos componentes aleatórios de $\mathbf{U}_i$. Supondo a existência de $\mathbb{E}[\exp\{z'_{ij}\mathbf{U}_i\}]$ e considerando $\text{Cov}(\mathbf{U}_i) = \mathbf{D}_i$,

$$\mathbb{E}(Y_{ij}) = \exp\{x'_{ij}\beta^{SS} + z'_{ij}\mathbf{D}_i z_{ij}/2\}, \quad (6.33)$$

encontrando semelhança com (6.29). Com relação à média marginal (modelo PA), temos que

$$\mathbb{E}(Y_{ij}) = \exp\{x'_{ij}\beta^{PA}\}, \quad (6.34)$$

dado que os efeitos aleatórios do modelo PA não são estimados do mesmo modo como nos modelos SS.

6.3.1 A Análise do Modelo Pós-teste

Young et al. (2007) analisaram dados referentes a um programa de prevenção de bebidas alcoólicas e de acidentes de carros, patrocinado pelo Escritório de Justiça Juvenil e Prevenção de Delinquência dos EUA. Em alguns estados, 52 comunidades que sofreram intervenção foram comparadas com outras 52 comunidades-controle. Indivíduos de 16 a 20 anos de idade participaram do estudo e foram entrevistados antes (pré-intervenção) e depois (pós-intervenção). As entrevistas versavam sobre várias questões, tais como experimentar consequências violentas com bebidas, compulsão por bebidas, adoção de comportamentos de risco envolvendo beber e dirigir etc. Os dados referem-se a 1868 jovens de 104 comunidades.

O modelo marginal (PA) pós-teste proposto pelos autores é dado por

$$\log(\mu_{ij}) = \beta_0^{PA} + \beta_1^{PA} TRT_i, \quad (6.35)$$

$$V(Y_{ij}) = \mathbb{E}[Y_{ij}] \phi^{PA} = \mu_{ij} \phi^{PA}, \quad (6.36)$$

sendo a variável resposta Y_{ij} representante da resposta do j -ésimo jovem da i -ésima comunidade sobre o número de passeios de carro nos últimos 30 dias com um motorista que estava bêbado (RIDES30).

A variável indicadora $TRT_i = 1$, se a comunidade do indivíduo respondente sofreu intervenção, e $TRT_i = 0$, caso contrário.

A estrutura de correlação *exchangeable* foi adotada, de forma que $\text{Corr}(Y_{ij}, Y_{ij'}) = \alpha$.

Em um modelo de Poisson, a média é igual à variância, de modo que $\phi^{PA} = 1$ em (6.36). Mas, a possibilidade de existência de sobredispersão nos dados, isto é, $\phi^{PA} > 1$, também é factível.

Aliás, McCulloch e Searle (2001, p. 223) demonstraram que um simples modelo marginal (obtido a partir de um modelo condicional com resposta Poisson e função de ligação *log*) de intercepto aleatório contém, por si só, sobredispersão. A demonstração desse resultado pode ser encontrada no Apêndice deste capítulo (Seção 6.6).

O modelo GLMM (SS) é dado por

$$\log(v_{ij}) = (\beta_0^{SS} + b_i) + \beta_1^{SS} TRT_i, \quad (6.37)$$

$$v_{ij} = \mathbb{E}[Y_{ij}|b_i], \quad (6.38)$$

$$b_i \sim N(0, \sigma_1^2), \quad (6.39)$$

com b_i representando o desvio aleatório da comunidade.

De forma a poder realizar uma comparação entre esses modelos, a média marginal deverá ser obtida, a partir da média condicional e escrita em termos de $\log(\mu_{ij})$. Inicialmente, devemos encontrar a média condicional.

$$\begin{aligned}
 \mu_{ij} &= \mathbb{E} \left[\mathbb{E}[Y_{ij}|b_i] \right] \\
 &= \mathbb{E}[v_{ij}] \\
 &= \mathbb{E}[\exp\{\beta_0^{SS} + b_i + \beta_1^{SS}TRT_i\}] \\
 &= \exp\{\beta_0^{SS} + \beta_1^{SS}TRT_i\} \mathbb{E}[\exp\{b_i\}] \\
 &= \omega_1 \exp\{\beta_0^{SS} + \beta_1^{SS}TRT_i\}, \tag{6.40}
 \end{aligned}$$

sendo que $\omega_1 = \exp\{\sigma_1^2/2\}$. Logo, ao tomarmos o logaritmo de ambos os lados, teremos

$$\log(\mu_{ij}) = (\beta_0^{SS} + \sigma_1^2/2) + \beta_1^{SS}TRT_i. \tag{6.41}$$

Nesse caso, já é possível fazer uma comparação entre (6.35) e (6.37). Por meio de (6.35) e (6.39), podemos concluir que a diferença dos interceptos entre os modelos (6.35) e (6.37) diferem, precisamente, por $\sigma_1^2/2$. Assim, o efeito medido pelo modelo PA (ao menos para o intercepto aleatório desse modelo de regressão de Poisson) é maior que o efeito medido pelo modelo SS, em $\sigma_1^2/2$, exatamente o resultado contrário daquele encontrado por Neuhaus et al. (1991), em um contexto de regressão logística, a saber, que, ao menos para um β pequeno, $|\beta_{PA}(X)| < |\beta|$, de modo que o efeito medido pelo modelo PA é menor que o efeito medido pelo modelo SS.

Entretanto, de acordo com os mesmos modelos (6.35) e (6.37), os coeficientes representados por β_1 não apresentaram diferenças, de modo que é possível considerar $\beta_1 = \beta_1^{PA} = \beta_1^{SS}$.

Young et al. (2007) ressaltam, ainda, que, ao incluir b_i no modelo, está-se explicitamente induzindo correlação entre as medidas repetidas dentro das comunidades, implicando no fenômeno de sobredispersão, isto é, $V(Y_{ij}) > E(Y_{ij})$. Somente quando $\sigma_1^2 = 0$, os interceptos aleatórios dos modelos PA e SS serão os mesmos.

As estimativas dos parâmetros dos modelos SS e PA referentes à aplicação do modelo Pós-Teste encontram-se na Tabela 6.5.

Como é possível notar, as estimativas dos modelos GEE ficaram bem próximas, independentemente de adotar uma estrutura de correlação, no caso, *exchangeable*, ou de optar por uma estrutura de *independência*.

Os coeficientes de tratamento $\beta_1 = \beta_1^{PA} = \beta_1^{SS}$ não se mostraram estatisticamente significantes. Mas, após a remoção de observações *outliers* (resíduos padronizados maiores que 2), todos passaram a apresentar significância estatística.

O fato de todas as estimativas serem negativas sugere que o efeito percebido nas comunidades que receberam o tratamento (intervenção) é, em média, de redução no número de passeios de carro nos últimos 30 dias na presença de motoristas que bebiam.

Apesar da proposição, por parte dos autores, de $\beta_1 = \beta_1^{PA} = \beta_1^{SS}$, ao menos com $N=1868$, as estimativas apresentaram uma diferença razoável (entre PA e SS). Mas com a remoção dos *outliers*, essa relativa diferença praticamente desapareceu.

Tabela 6.5 – Estimativas de PA e SS para modelo Pós-Teste

Method	Parameter	All available data ($N = 1868$)			Excluding outliers ($N = 1827$)		
		Estimate	SE ^a	Z	Estimate	SE ^a	Z
GEE-ind	Intercept	-0.718	0.146	-4.93	-1.297	0.092	-14.07
	Treatment	-0.286	0.206	-1.39	-0.287	0.124	-2.31
	ϕ	6.746			1.755		
GEE-exch	Intercept	-0.712	0.136	-5.25	-1.298	0.084	-15.67
	Treatment	-0.289	0.207	-1.39	-0.286	0.127	-2.26
	ϕ	6.718			1.756		
GLMM	Intercept	-1.058	0.126	-8.37	-1.383	0.092	-15.05
	Treatment	-0.230	0.178	-1.29	-0.288	0.130	-2.21
	σ_1^2	0.621	0.117	5.31	0.184	0.063	2.91

^aEmpirical sandwich standard errors are reported for GEE with working independence correlation. Otherwise, model-based standard errors are reported.

Fonte: Young et al. (2007, p. 176).

Ainda sobre os modelos GEE, Young et al. (2007) relataram uma elevada estimativa de sobredispersão, com ϕ^{PA} bem maior que 1. De modo similar, ao remover *outliers*, essas estimativas diminuíram consideravelmente, tendo ficado, ambas, ainda acima de 1, mas, inferiores a 2.

Esse efeito dos *outliers* também foi percebido pelos autores acerca das correlações entre as medidas repetidas. Fazendo uma separação entre as UEs tratadas e não tratadas, temos que, para o modelo SS, $\alpha_c = \text{Corr}(Y_{ij}, Y_{ij'} | TRT_i = 0)$ e $\alpha_i = \text{Corr}(Y_{ij}, Y_{ij'} | TRT_i = 1)$. Para o modelo PA, a estrutura de correlação é a mesma (*exchangeable*). Assim, por meio da Tabela 6.6, é possível notar que, com a remoção dos *outliers*, as medidas α_c e α_i também diminuíram consideravelmente.

Tabela 6.6 – Estimativas das correlações nos modelos PA e SS

Correlation	All available data		Excluding outliers	
	GEE-exch	GLMM (2)	GEE-exch	GLMM (2)
α_c	0.015	0.290	-0.001	0.052
α_i	0.015	0.245	-0.001	0.040

Fonte: Young et al. (2007, p. 176).

6.3.2 A Análise do Modelo Pré-teste/Pós-teste

Já nesse tipo de modelo, Young et al. (2007) introduzem uma variável $TIME_t$ para viabilizar a análise pré-intervenção e pós-intervenção, de modo que $TIME_t = 0$, se pré-teste ($t = 1$), e $TIME_t = 1$, se pós-teste ($t = 2$).

O efeito de interesse por parte dos autores reside na análise do coeficiente da interação ($TRT_i \times TIME_t$), ou seja, em β_3^{PA} , como forma de se medir o próprio efeito do tratamento ao longo do tempo. O modelo marginal (PA) é dado pelas duas equações a seguir:

$$\log(\mu_{ijt}) = \beta_0^{PA} + \beta_1^{PA}TRT_i + \beta_2^{PA}TIME_t + \beta_3^{PA}(TRT_i \times TIME_t) \quad (6.42)$$

$$V(Y_{ijt}) = \mu_{ijt}\phi^{PA}, \quad (6.43)$$

com $\mu_{ijt} = \mathbb{E}[Y_{ijt}]$ e $Y_{ijt} = RIDES30$ para o i -ésimo jovem pertencente à j -ésima comunidade em $TIME_t = 0$ ($t = 1$) ou $TIME_t = 1$ ($t = 2$).

Já o modelo condicional (SS) é dado por

$$\begin{aligned} \log(v_{ijt}) = & (\beta_0^{SS} + b_i) + \beta_1^{SS}TRT_i + (\beta_2^{SS} + u_i)TIME_t \\ & + \beta_3^{SS}(TRT_i \times TIME_t) \end{aligned} \quad (6.44)$$

com $v_{ijt} = \mathbb{E}[Y_{ijt}|b_i, u_i]$, sendo b_i o efeito aleatório da comunidade e u_i o desvio de *community* $\times$ *time* do efeito *específico-comunidade*.

Young et al. (2007) supõem, ainda, b_i e u_i independentes e distribuídos segundo

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{u} \end{pmatrix} \sim NMV\left(\mathbf{0}, \begin{pmatrix} I\sigma_1^2 & 0 \\ 0 & I\sigma_2^2 \end{pmatrix}\right), \quad (6.45)$$

com $\mathbf{b} = (b_1, \dots, b_K)$ e $\mathbf{u} = (u_1, \dots, u_K)$ vetores de efeitos aleatórios de ordem $K \times 1$. Por um lado, b_i permite que o intercepto varie pelas comunidades. Por outro, u_i refere-se a um termo que permite que o coeficiente β_2 , devido ao tempo, também possa variar pelas comunidades.

Além disso, Young et al. (2007) assumiram independência entre os efeitos aleatórios, ou seja, $Cov(b_i, u_i) = 0$.

De modo semelhante à configuração pós-teste, para poder realizar uma comparação entre os modelos PA e SS na configuração pré-teste/pós-teste, a média marginal também deverá ser obtida a partir da média condicional.

Com base nas suposições de normalidade e independência dos efeitos aleatórios, a média marginal μ_{ijt} , a partir do modelo de média condicional $Y_{ijt}|b_i, u_i$, será dada por

$$\begin{aligned}\mu_{ijt} &= \mathbb{E} \left[\mathbb{E}[Y_{ijt}|b_i, u_i] \right] \\ &= \mathbb{E}[v_{ijt}] \\ &= \exp\{\beta_0^{SS} + \beta_1^{SS}TRT_i + \beta_2^{SS}TIME_t \\ &\quad + \beta_3^{PA}(TRT_i \times TIME_t)\} \mathbb{E}[\exp\{b_i\}] \mathbb{E}[\exp\{u_i \times TIME_t\}],\end{aligned}\tag{6.46}$$

de modo que

$$\mathbb{E}[\exp\{b_i\}] = \exp\{\sigma_1^2/2\} \equiv \omega_1,\tag{6.47}$$

$$\mathbb{E}[\exp\{u_i\}] = \exp\{\sigma_2^2/2\} \equiv \omega_2\tag{6.48}$$

e

$$\mathbb{E}[\exp\{u_i \times TIME_t\}] = \begin{cases} 1, & \text{em } t = 1 \\ \omega_2, & \text{em } t = 2. \end{cases}\tag{6.49}$$

Logo, ao realizarmos essas três substituições e tomarmos o logaritmo de ambos os lados, teremos o seguinte modelo para a média marginal:

$$\begin{aligned}\log(\mu_{ijt}) &= \left(\beta_0^{SS} + \frac{\sigma_1^2}{2} \right) + \beta_1^{SS}TRT_i + \left(\beta_2^{SS} + \frac{\sigma_2^2}{2} \right) TIME_t \\ &\quad + \beta_3^{SS}(TRT_i \times TIME_t).\end{aligned}\tag{6.50}$$

Nesse caso, já é possível fazer uma comparação entre os modelos (6.42) e (6.44). Por meio de (6.50), é possível notar que os efeitos medidos por β_0^{PA} e β_2^{PA} (do modelo PA) são menores quando comparados com os mesmos efeitos medidos pelo modelo SS (β_0^{SS} e β_2^{SS}), enquanto que, de outro modo, as estimativas de β_1 e β_3 podem ser comparadas diretamente entre os modelos PA e SS, ora descritos em (6.42) e (6.44).

Também de modo similar ao modelo pós-teste, Young et al. (2007) chamam atenção para o fato de que, somente quando $\sigma_1^2 = \sigma_2^2 = 0$, os modelos seriam equivalentes (PA e SS). Se esse não for o caso, o modelo (6.44) conterá sobredispersão.

Para a estimação dos modelos pré-teste/pós-teste, dados de 3.621 jovens foram utilizados. O tamanho das UEs (jovens observados nos dois períodos de tempo por comunidade) tem variado de 27 a 55, com média de 35.

Nessa configuração de modelo pré-teste/pós-teste, a variável resposta RIDES30 apresentou a seguinte distribuição: 0 (82%), 1 (8%), 2 (5%), 3 (2%) e de 4 a 30 (3%).

A Tabela 6.7 apresenta as estimativas para o modelo GEE, com matrizes de correlação: *independente*, *exchangeable* e *baseada em planejamento* (*design-based*).

Os autores Young et al. (2007) descrevem a matriz de correlação com as seguintes características: seja a matriz de correlação para $\mathbf{Y}_i$ denotada por $R_i(\alpha)$, com $\alpha = \{\alpha_0, \alpha_1\}$, tal que $\alpha_0 = \text{Corr}(\mathbf{Y}_{ijt}, \mathbf{Y}_{ij't})$ é a correlação entre dois indivíduos distintos dentro de uma mesma comunidade e em um mesmo ponto do tempo e $\alpha_1 = \text{Corr}(\mathbf{Y}_{ijt}, \mathbf{Y}_{ij't'})$ é a correlação entre dois indivíduos distintos em uma mesma comunidade e em dois pontos distintos do tempo. Logo, a estrutura de correlação baseada em planejamento será dada por

$$R_i(\alpha_0, \alpha_1) = \begin{pmatrix} R_{i1} & R_{i12} \\ R'_{i12} & R_{i2} \end{pmatrix}, \quad (6.51)$$

com a observação de que esse tipo de matriz permite acomodar mais de dois períodos.

Tabela 6.7 – Estimativas de modelos PA

Correlation structure	Variable	All available data ($N = 3621$)			Outliers excluded ($N = 3532$)		
		Estimate	SE ^a	Z	Estimate	SE ^a	Z
Independence	Intercept	-0.516	0.134	-3.86	-1.222	0.084	-14.50
	Treatment	-0.079	0.176	-0.45	0.023	0.117	0.20
	Time	-0.202	0.178	-1.14	-0.075	0.117	-0.64
	Trt $\times$ Time	-0.207	0.267	-0.77	-0.310	0.167	-1.85
	ϕ	7.164			1.795		
Exchangeable	Intercept	-0.514	0.122	-4.23	-1.222	0.084	-14.50
	Treatment	-0.080	0.176	-0.46	-0.023	0.119	0.19
	Time	-0.202	0.171	-1.18	-0.075	0.119	-0.63
	Trt $\times$ Time	-0.206	0.255	-0.81	-0.310	0.176	-1.76
	ϕ	7.157			1.795		
Design-based	Intercept	-0.516	0.124	-4.15	-1.222	0.084	-14.50
	Treatment	-0.079	0.180	-0.44	0.023	0.119	0.19
	Time	-0.200	0.179	-1.12	-0.075	0.117	-0.64
	Trt $\times$ Time	-0.208	0.266	-0.78	-0.310	0.172	-1.80
	ϕ	7.158			1.794		

^aEmpirical sandwich standard errors are reported for GEE with working independence correlation. Otherwise, model-based standard errors are reported. Exchangeable correlation estimate for GEE is $\alpha = 0.006$ based upon all available data and $\alpha = 0.0001$ when outliers are removed. Design-based GEE correlation estimates are ($\alpha_0 = 0.008$, $\alpha_1 = 0.003$) based upon all available data and ($\alpha_0 = 0.0001$, $\alpha_1 = 0.002$) when outliers are removed.

Fonte: Young et al. (2007, p. 177).

O efeito de interesse ($TRT_i \times TIME_t$) não se mostrou estatisticamente significativo, mesmo com a remoção de *outliers* baseados em resíduos padronizados com valores maiores que 2.

Os três modelos PA (com estruturas de correlação *independence*, *exchangeable* e *design-based*) apresentaram elevada sobredispersão (7,184, 7,157 e 7,158), sendo que, após a remoção dos *outliers*, todas essas medidas diminuíram consideravelmente (1,795, 1,795 e 1,794).

As estimativas para os modelos SS encontram-se na Tabela 6.8. Neste caso, além de um modelo GLMM com a suposição de independência entre os efeitos aleatórios, isto é, $Cov(b_i, u_i) = 0$, também foi estimado um modelo GLMM com estrutura de covariância *unstructured* entre os efeitos aleatórios, isto é, $Cov(b_i, u_i) = \sigma_{12}$.

Tabela 6.8 – Estimativas de modelos SS

Variance components	Variable	All available data ($N = 3621$)			Outliers excluded ($N = 3532$)		
		Estimate	SE	Z	Estimate	SE	Z
Independence model (4)	Intercept	-0.788	0.109	-7.24	-1.265	0.076	-16.63
	Treatment	-0.011	0.152	-0.07	0.029	0.106	0.28
	Time	-0.350	0.158	-2.21	-0.133	0.109	-1.23
	Trt $\times$ Time	-0.184	0.224	-0.82	-0.312	0.154	-2.03
	σ_1^2	0.461	0.083	5.55	0.081	0.031	2.59
	σ_2^2	0.937	0.184	5.09	0.148	0.067	2.21
Unstructured	Intercept	-0.844	0.121	-6.97	-1.303	0.088	-14.86
	Treatment	-0.007	0.169	-0.04	0.031	0.120	0.26
	Time	-0.215	0.173	-1.25	-0.081	0.124	-0.65
	Trt $\times$ Time	-0.222	0.242	-0.92	-0.318	0.172	-1.85
	σ_1^2	0.585	0.107	5.48	0.155	0.054	2.89
	σ_2^2	1.162	0.215	5.41	0.299	0.109	2.75
	σ_{12}	-0.564	0.128	-4.42	-0.135	0.065	-2.09

Fonte: Young et al. (2007, p. 178).

Com relação ao coeficiente de interesse, β_2^{SS} , as estatísticas Z resultaram em -0,82 e -0,92 nos modelos GLMM *independence* e *unstructured*, respectivamente. Isso significa que os efeitos da intervenção ao longo do tempo não se mostraram significantes. Young et al. (2007) sugerem que o erro padrão do parâmetro pode ter sido subestimado em razão de uma importante fonte de variação tal como a covariância entre interceptos e coeficientes.

Quando *outliers* são removidos, o parâmetro que mede o efeito da intervenção ao longo do tempo torna-se significativo (no modelo GLMM *independence*) com *p-valor* de 0,04 (informado pelos próprios autores). Já com relação ao modelo GLMM *unstructured*, o efeito da intervenção permanece sem significância estatística, com *p-valor* de 0,07 (estimativa também informada pelos autores).

Segundo ainda os autores, quando os *outliers* são removidos, os parâmetros de intervenção dos modelos GEE e GLMM tendem a ficar bem próximos (resultado já antecipado, uma vez que era esperado que $\beta_3^{PA} = \beta_3^{SS}$).

A relação exata entre os parâmetros SS e PA, em ambas as configurações (pós-teste e pré-teste/pós-Teste), pode ser analisada, com mais detalhes, na Tabela 6.9, também fornecida por Young et al. (2007).

Nesta Tabela 6.9, todos os valores das colunas PA (“All available data” e “Excluding outliers”) são de estimativas que se encontram nas Tabelas 6.5, 6.7 e 6.8 e, portanto, não se referem a aproximações.

Os valores das colunas SS, especificamente nas linhas: i) $\beta_1^{PA} = \beta_1^{SS}$ (“Posttest-only”), ii) $\beta_1^{PA} = \beta_1^{SS}$ (“Pretest-Posttest”) e iii) $\beta_3^{PA} = \beta_3^{SS}$ (“Pretest-Posttest”) também são estimativas dos modelos e, portanto, também não são aproximações.

As aproximações encontram-se somente nas colunas SS (“All available data” e “Excluding outliers”), precisamente nas linhas: i) $\beta_0^{PA} = \beta_0^{SS} + \sigma_1^2/2$ (“Posttest-only”), ii) $\beta_0^{PA} = \beta_0^{SS} + \sigma_1^2/2$ (“Pretest-Posttest”) e iii) $\beta_2^{PA} = \beta_2^{SS} + \sigma_2^2/2$ (“Pretest-Posttest”).

Tabela 6.9 – Relação entre os parâmetros dos modelos PA e SS

Equation ^a	All available data		Excluding outliers	
	PA	SS	PA	SS
Posttest-only				
$\beta_0^{PA} = \beta_0^{SS} + \sigma_1^2/2$	−0.712	−0.748	−1.298	−1.291
$\beta_1^{PA} = \beta_1^{SS}$	−0.289	−0.230	−0.286	−0.288
Pretest–Posttest				
$\beta_0^{PA} = \beta_0^{SS} + \sigma_1^2/2$	−0.516	−0.558	−1.222	−1.225
$\beta_1^{PA} = \beta_1^{SS}$	−0.079	−0.011	0.023	0.029
$\beta_2^{PA} = \beta_2^{SS} + \sigma_2^2/2$	−0.200	0.119	−0.075	−0.059
$\beta_3^{PA} = \beta_3^{SS}$	−0.208	−0.184	−0.310	−0.312

^aColumn entries under PA are estimates for the left side of the equation; entries under SS are estimates for the right side of the equation. PA posttest-only results are based upon GEE with exchangeable correlation structure; pretest–posttest results are based upon the design-based correlation structure. SS model results are based upon models (2) and (4), respectively.

Fonte: Young et al. (2007, p. 179).

As aproximações para o modelo pós-teste baseadas nas estimativas do modelo SS (tendo o modelo GLMM-exch, como referência para comparação pelos próprios autores) são dadas por:

- Para o intercepto (“All available data”): $\beta_0^{PA} = \beta_0^{SS} + \frac{\sigma_1^2}{2} = -1,058 + \frac{0,621}{2} = -0,7475 \approx -0,748$;
- Para o intercepto (“Excluding outliers”): $\beta_0^{PA} = \beta_0^{SS} + \frac{\sigma_1^2}{2} = -1,383 + \frac{0,184}{2} = -1,291$.

Dentre os modelos pós-teste (com e sem remoção de outliers), o modelo com remoção de *outliers* (“Excluding outliers”) apresentou uma melhor aproximação com diferença (em valor absoluto) de 0,007, tendo o modelo GEE-exch como referência para comparação. A aproximação do modelo sem remoção de *outliers* teve diferença, em valor absoluto, de 0,036.

Para as aproximações para o modelo pré-teste/pós-teste, as estimativas do modelo SS (GLMM – *Design based*, como referência adotada pelos próprios autores) são dadas por:

- Para o intercepto (“All available data”): $\beta_0^{PA} = \beta_0^{SS} + \frac{\sigma_1^2}{2} = -0,788 + \frac{0,461}{2} = -0,5575 \approx -0,558$.
- Para o intercepto (“Excluding outliers”): $\beta_0^{PA} = \beta_0^{SS} + \frac{\sigma_1^2}{2} = -1,265 + \frac{0,081}{2} = -1,2245 \approx -1,225$.
- Para o parâmetro de intervenção (“All available data”): $\beta_2^{PA} = \beta_2^{SS} + \frac{\sigma_2^2}{2} = -0,350 + \frac{0,937}{2} = 0,1185 \approx 0,119$.
- Para o parâmetro de intervenção (“Excluding outliers”): $\beta_2^{PA} = \beta_2^{SS} + \frac{\sigma_2^2}{2} = -0,133 + \frac{0,148}{2} = -0,059$.

Dentre os modelos pré-teste/pós-teste, as melhores aproximações também ocorreram quando os *outliers* foram removidos (“Excluding outliers”), seja para o intercepto (diferença de 0,003 em valor absoluto), seja para o parâmetro da intervenção (diferença de 0,016, em valor absoluto).

As correlações intraclasse dos modelos PA e SS podem ser analisadas por meio da Tabela 6.10. No caso, as correlações estimadas pelo modelo GLMM são bem maiores, só diminuindo quando removidos os *outliers*. Mas, ainda assim, as correlações medidas pelo modelo GLMM mostraram-se superiores quando comparadas com as correlações obtidas pelo modelo GEE-*Design based*.

Tabela 6.10 – Correlações intraclasse entre modelos GEE e GLMM

Correlation	All available data		Excluding outliers	
	GEE-design	GLMM	GEE-design	GLMM
$\alpha_{c,11}$	0.008	0.370	0.0001	0.039
$\alpha_{c,12}$	0.003	0.251	0.002	0.046
$\alpha_{c,22}$	0.008	0.885	0.0001	0.163
$\alpha_{i,11}$	0.008	0.368	0.0001	0.040
$\alpha_{i,12}$	0.003	0.247	0.002	0.041
$\alpha_{i,22}$	0.008	0.864	0.0001	0.128

^aGEE results are based upon the design-based correlation structure. GLMM results are based upon maximum likelihood estimates for model (4).

Fonte: Young et al. (2007, p. 179).

De um modo geral, as estimativas tanto dos modelos GEE (com procedimentos robustos) quanto dos modelos GLMM (método da máxima verossimilhança) mostraram-se sensíveis à influência de observações *outliers*. Essa sensibilidade também foi constatada tanto nas estimativas das correlações intraclasse quanto nas estimativas de sobredispersão.

6.4 Os modelos PA e SS – Regressão de Poisson (parte II)

Zhang, Yu, Feng, Gunzler, Wu e Tu (2012) também produziram alguns resultados importantes ao analisarem, de um ângulo diferente, a relação existente entre os modelos SS e PA, em um contexto de dados longitudinais. No caso, os autores conduziram uma análise detalhada sobre os efeitos da sobredispersão implicada no modelo marginal sobre as probabilidades de erro tipo I (com relação aos testes de hipóteses acerca das estimativas dos parâmetros).

Além disso, Zhang et al. (2012) chamam atenção para o fato de que, em larga medida, as suposições implícitas nas técnicas de inferência para essas classes de modelos repousam na premissa de grandes amostras, apesar dos autores realizarem suas análises com dados simulados e reais, de pequeno e médio portes. Todas as análises foram realizadas com o software R (bibliotecas *geepack* e *lme4*).

Zhang et al. (2012), logo na introdução, reconhecem a literatura recentemente desenvolvida acerca desse tema ao citar dois livros (Diggle, Heagerty, Liang e Zeger (2002) e

Venables e Ripley (2008)) e três artigos (Young et al. (2007), Zeger et al. (1988) e Ritz e Spiegelman (2004)), sendo que este último artigo será objeto de análise ainda neste capítulo.

Os autores introduzem um modelo clássico de regressão de Poisson e paulatinamente vão sofisticando-o até alcançar a forma desejada. Inicialmente, consideram um modelo clássico de Poisson loglinear para análise de dados de corte transversal com n UEs. Definem y_i como uma resposta (de contagem) e $\mathbf{x}_i$, um vetor de preditoras (ou covariáveis) da i -ésima UE, com $i = 1, \dots, n$. Em outros termos,

$$y_i | \mathbf{x}_i \sim \text{Poisson}(\mu_i), \quad (6.52)$$

$$\mu_i = \mathbb{E}(y_i | \mathbf{x}_i) \quad (6.53)$$

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (6.54)$$

sendo y_i , dado $\mathbf{x}_i$, independentes.

Ainda que o modelo descrito pelas equações (6.52), (6.53) e (6.54) explicitamente não leve em consideração a possibilidade de modelar dados longitudinais, a suposição convencional da distribuição de Poisson, a saber, de que $\mathbb{E}(y_i | \mathbf{x}_i) = \text{Var}(y_i | \mathbf{x}_i)$, ainda está presente.

Ao ocorrer o fenômeno de sobredispersão, os autores chamam atenção para o fato de que estimativas de máxima verossimilhança para um modelo de Poisson não oferecem inferências válidas (erros padrões artificialmente diminuídos produzindo uma quantidade maior de parâmetros estatisticamente significantes do que efetivamente seria o caso).

Zhang et al. (2012) sugerem duas soluções para lidar com o fenômeno da sobredispersão:

- 1) Adotar a distribuição Binomial Negativa para a variável resposta, sendo uma distribuição também de contagem, com média igual à da distribuição de Poisson, mas com variância maior, tornando possível a acomodação de uma maior variabilidade dos dados.
- 2) Usar o estimador de robusto “sandwich” SVE (*Sandwich Variance Estimator*), pelo procedimento GEE.

Ao adotar a segunda solução, a distribuição de Poisson (6.52), implicada pelo modelo clássico de Poisson loglinear para análise de dados de corte transversal descrito em (6.52), (6.53) e (6.54), é descartada, de modo que o modelo, segundo Zhang et al. (2012), passaria a ser descrito somente por (6.53) e (6.54), também chamado pelos autores de *distribution-free log-linear model*.

Ao inserir o efeito do tempo t , isto é, a possibilidade de cada uma das UEs produzir m respostas ao longo do tempo, teremos y_{it} e $\mathbf{x}_{it}^T$, como a variável resposta de contagem e um

vetor de preditoras/covariáveis, respectivamente, para a i -ésima UE no instante t do tempo, com $i = 1, \dots, n$ e $1 \leq t \leq m$.

Assim, Zhang et al. (2012) qualificam o modelo GEE como aquele a partir do qual a correlação das respostas y_{it} é levada em consideração em cada um dos instantes t do tempo, em um contexto de *distribution-free log-linear model*. Assim, para cada instante t do tempo,

$$\mu_{it} = \mathbb{E}(y_{it} | \mathbf{x}_{it}), \quad (6.55)$$

$$\log(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}, \quad (6.56)$$

para $i = 1, \dots, n$ e $1 \leq t \leq m$.

Por fim, Zhang et al. (2012) apresentam o seguinte modelo GLMM:

$$y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{b}_i \sim \text{Poisson}(\tilde{\mu}_{it}), \quad (6.57)$$

$$\log(\tilde{\mu}_{it}) = \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} + \mathbf{z}_{it}^T \mathbf{b}_i \quad (6.58)$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \Sigma_b), \quad (6.59)$$

com $i = 1, \dots, n$ e $1 \leq t \leq m$.

Note que $\tilde{\mu}_{it}$ e $\tilde{\boldsymbol{\beta}}$ estão associados ao modelo GLMM, representado pelas equações (6.57), (6.58) e (6.59), enquanto que μ_{it} e $\boldsymbol{\beta}$ estão associados ao modelo GEE, representado pelas equações (6.55) e (6.56).

Zhang et al. (2012) lembram que o fato de $\mathbf{b}_i$ ser variável latente torna bastante difícil a validação empírica de um modelo matemático para sua distribuição. No entanto, a introdução de $\mathbf{b}_i$ altera a suposição de Poisson, uma vez que

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}) = \exp \left\{ \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} + \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \quad (6.60)$$

$$\text{Var}(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}) \geq \mathbb{E}(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}), \quad (6.61)$$

tendo (6.61) sido demonstrado no apêndice deste capítulo (Seção 6.6).

Assim, somente se $\Sigma_b = 0$, a igualdade estaria assegurada, ou seja, se e somente se os y_{it} 's não fossem correlacionados. Uma vez que, na maioria dos estudos, os y_{it} 's são correlacionados, a igualdade entre $\text{Var}(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it})$ e $\mathbb{E}(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it})$ não se verifica, evidenciando o fenômeno de sobredispersão, de modo que o modelo marginal, representado pelas equações (6.60) e (6.61), não possui distribuição de Poisson.

De forma a tornar comparável o modelo marginal, representado pelas equações (6.60) e (6.61), com o modelo GEE, representado pelas equações (6.55) e (6.56), os autores notaram que

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}) = \mathbb{E} \left[\exp \left\{ \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} + \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \mid \mathbf{x}_{it} \right] \quad (6.62)$$

$$= \mathbb{E} \left[\exp \left\{ \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \mid \mathbf{x}_{it} \right] \exp \{ \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} \},$$

com a observação de uma forma fechada para o termo da esperança condicional do lado direito de (6.62) seria de difícil obtenção. Nesse caso, Zhang et al. (2012) destacaram dois casos especiais que permitiriam a obtenção da forma fechada dessa esperança condicional:

- 1) Se $\mathbf{z}_{it}$ for estocasticamente independente de $\mathbf{x}_{it}$, então, $\mathbb{E} \left[\exp \left\{ \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \mid \mathbf{x}_{it} \right] =$

$$\mathbb{E} \left[\exp \left\{ \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \right] = \exp \{ \gamma_0 \}$$
 e a expressão (6.62) reduz-se a

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}) = \exp \{ \gamma_0 + \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} \}. \quad (6.63)$$

Assim, a não ser pelo termo do intercepto, os parâmetros contidos em $\tilde{\boldsymbol{\beta}}$ conterão as mesmas interpretações admitidas para $\boldsymbol{\beta}$, no modelo GEE, em (6.55) e (6.56).

- 2) Se $\mathbf{z}_{it}$ for um subvetor de $\mathbf{x}_{it}$. Para fins de simplificação, considere o caso em que $\mathbf{x}_{it}$ não contenha o intercepto, de modo que $\mathbf{x}_{it} = (\mathbf{z}_{it}^T, \mathbf{w}_{it}^T)^T$. Procede-se de modo semelhante com o vetor paramétrico, de forma que $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_z^T, \tilde{\boldsymbol{\beta}}_w^T)^T$. Então, $\mathbb{E} \left[\exp \left\{ \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \mid \mathbf{x}_{it} \right] = \exp \left\{ \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\}$, de modo que a expressão (6.62) reduz-se a

$$\begin{aligned} \mathbb{E}(y_{it} | \mathbf{x}_{it}) &= \exp \left\{ \mathbf{z}_{it}^T \tilde{\boldsymbol{\beta}}_z + \mathbf{w}_{it}^T \tilde{\boldsymbol{\beta}}_w + \frac{1}{2} \mathbf{z}_{it}^T \Sigma_b \mathbf{z}_{it} \right\} \\ &= \exp \left\{ \mathbf{w}_{it}^T \tilde{\boldsymbol{\beta}}_w + \mathbf{z}_{it}^T \left(\tilde{\boldsymbol{\beta}}_z + \frac{1}{2} \Sigma_b \mathbf{z}_{it} \right) \right\}. \end{aligned} \quad (6.64)$$

Logo, $\tilde{\boldsymbol{\beta}}_w$ terá uma interpretação similar com aqueles parâmetros correspondentes ao modelo GEE, no entanto, a interpretação de $\tilde{\boldsymbol{\beta}}_z$ é modificada uma vez que sofre um deslocamento proporcional a $\frac{1}{2} \Sigma_b \mathbf{z}_{it}$.

Além disso, os autores notaram que a(s) variável(is) que estiver(em) em $\mathbf{x}_{it}$ e que se sobrepo(m) em $\mathbf{z}_{it}$ pode(m), eventualmente, ser estatisticamente significante(s) em um modelo, mas não em outro.

Uma terceira possibilidade abordada em Zhang et al. (2012) consiste em $\mathbf{z}_{it}$ estar correlacionado com $\mathbf{x}_{it}$ sem, contudo, ser um subvetor de $\mathbf{x}_{it}$. Nesse caso, os autores assinalam que modelos com respostas de Poisson, GLMM, loglineares (equações (6.57), (6.58) e (6.59)) são capazes de produzir conclusões diferentes sendo, eventualmente, impossível de se examinar a relação entre $\boldsymbol{\beta}$ e $\tilde{\boldsymbol{\beta}}$ devido ao maior grau de complexidade. Esse caso particular de correlação entre $\mathbf{z}_{it}$ e $\mathbf{x}_{it}$, sem que $\mathbf{z}_{it}$ seja necessariamente um subvetor de $\mathbf{x}_{it}$ também será investigado pelos autores, constituindo uma situação diferente daquelas analisadas neste capítulo.

Os testes de hipóteses a serem realizados pelos autores seguem a forma

$$\begin{cases} H_0: \boldsymbol{\beta}_1 = \mathbf{c} \\ H_1: \boldsymbol{\beta}_1 = \mathbf{d} \end{cases}$$

com $\mathbf{c} \neq \mathbf{d}$ refletindo constantes conhecidas e $\boldsymbol{\beta}_1$ representando alguns componentes ou mesmo um subvetor de $\boldsymbol{\beta}$.

As estatísticas de Wald são amplamente utilizadas para testar hipóteses tais como $H_0: \boldsymbol{\beta}_1 = \mathbf{c}$ contra $H_1: \boldsymbol{\beta}_1 = \mathbf{d} \neq \mathbf{c}$, apesar de serem anticonservativas, isto é, com probabilidades de erro tipo I infladas.

Uma alternativa, também, para as estatísticas Wald (e bastante utilizada) é o TRV (teste da razão de verossimilhanças), o qual, apesar de assintoticamente equivalente, supera, de acordo com os autores, em desempenho, a estatística Wald em situações de amostras pequenas.

Os autores apresentaram a estatística do teste de score e em seguida implementaram uma correção via T^2 de Hotelling. A mesma correção foi feita com a estatística do TRV, no intuito de melhorar o desempenho para amostras reduzidas.

6.4.1 Análise de Dados Simulados

A primeira análise realizada pelos autores consiste em uma base de dados simulada, com medidas repetidas em três pontos distintos do tempo, a partir do seguinte modelo GLMM:

$$y_{it} | \mathbf{x}_{it}, b_i \sim \text{Poisson}(\tilde{\mu}_{it}), \quad (6.65)$$

$$\log(\tilde{\mu}_{it}) = \tilde{\beta}_0 + x_i \tilde{\beta}_1 + z_i b_i, \quad (6.66)$$

$$b_i \sim N(0, \tau^2), \quad (6.67)$$

com $x_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, para $1 \leq i \leq n$ e $1 \leq t \leq 3$.

Os autores notaram, ainda, que quando τ^2 diminui, a correlação intraclass também diminui, de modo que y_{it} , condicionado em $\mathbf{x}_{it}$, aproxima-se de um modelo GEE

$$\mu_{it} = \mathbb{E}(y_{it} | \mathbf{x}_{it}), \quad (6.68)$$

$$\log(\mu_{it}) = \beta_0 + x_i \beta_1, \quad (6.69)$$

com $1 \leq i \leq n$ e $1 \leq t \leq 3$.

Com relação aos dois casos especiais mencionados anteriormente que permitem obter uma forma fechada para a expressão (6.62), bem como a possibilidade adicional da análise de uma relação mais complexa, Zhang et al. (2012) conseguiram colocar à prova essas três possibilidades ao caracterizar z_i em três situações específicas:

- a) $z_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ e estocasticamente independente de x_i ;

- b) $z_i = x_i$;
 c) $z_i \sim \chi_1^2 I_{\{x_i > 0\}}$.

Na primeira situação, z_i e x_i são independentes, apesar de exibirem a mesma distribuição de probabilidades de mesmo parâmetro. Na segunda situação, z_i é subconjunto de x_i , enquanto que, na terceira, z_i é correlacionado com x_i , sem que seja um subconjunto de x_i . Nessa terceira situação, se $x_i = 1$ (isto é, $x_i > 0$), z_i assume distribuição qui-quadrado (com 1 grau de liberdade). Do contrário, $z_i = 0$, se $x_i = 0$. Dessa maneira, garante-se que z_i não seja subconjunto de x_i mas, que, ao apresentarem certo padrão, estejam correlacionados.

É importante notar, também, que, z_i passa a admitir uma distribuição de probabilidades, o que não acontecia nos casos anteriores, de modo que as expressões (6.63) e (6.64), relativas às situações (a) e (b), respectivamente, sofrerão leves modificações.

Para a situação (a), temos que

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}) = \mathbb{E} \left[\exp \left\{ \tilde{\beta}_0 + x_i \tilde{\beta}_1 + \frac{1}{2} z_i \tau^2 \mid \mathbf{x}_{it} \right\} \right] = \exp \{ (\tilde{\beta}_0 + \gamma_0) x_i \tilde{\beta}_1 \}, \quad (6.70)$$

que, conforme já antecipado, $\tilde{\beta}_1$ terá a mesma interpretação que β_1 , do modelo GEE, enquanto que $\tilde{\beta}_0$ sofre um deslocamento de magnitude γ_0 .

Para a situação (b), temos que

$$\begin{aligned} \mathbb{E}(y_{it} | \mathbf{x}_{it}) &= \mathbb{E} \left[\exp \left\{ \tilde{\beta}_0 + x_i \tilde{\beta}_1 + \frac{1}{2} z_i \tau^2 \mid \mathbf{x}_{it} \right\} \right] \\ &= \exp \left\{ \tilde{\beta}_0 + x_i \left(\tilde{\beta}_1 + \frac{1}{2} \tau^2 \right) \right\}, \end{aligned} \quad (6.71)$$

tendo, agora, $\tilde{\beta}_1$ sido deslocado em $\delta_1 = \frac{1}{2} \tau^2$, quando comparado com β_1 correspondente ao modelo GEE. Para $\tilde{\beta}_0$, a interpretação é a mesma de β_0 , do modelo GEE.

A situação (c), em razão de sua complexidade, não possui uma forma fechada. Esse fato não impede a comparação das estimativas dos parâmetros de GLMM e GEE. O problema é que, comparativamente a (a) e (b), não é possível comparar as diferenças dos parâmetros em termos de uma relação matemática precisa. Nesse caso, Zhang et al. (2012) fizeram uma comparação pela diferença média entre as estimativas. As estimativas dos três modelos são apresentadas na Tabela 6.11.

Com relação à situação (a):

- Em $\tau^2 = 0,5$, o parâmetro $\tilde{\beta}_0$ sofreu um deslocamento da ordem de 0,13, enquanto que $\tilde{\beta}_1$ teve o mesmo valor estimado de β_1 , conforme esperado.
- Em $\tau^2 = 1$, o parâmetro $\tilde{\beta}_0$ sofreu um deslocamento maior (0,28), enquanto que $\tilde{\beta}_1$ teve praticamente o mesmo valor estimado de β_1 , com leve modificação.

Tabela 6.11 – Estimativas GEE e GLMM para três situações específicas de $\mathbf{z}_i$

Table 1. Comparison of parameter estimates between the GLMM and GEE models for three structures of $\mathbf{z}_i$, with averaged difference between the estimates (Ave. $\hat{\beta} - \tilde{\beta}$) of the GLMM and GEE obtained based on 1000 MC replications.

	$x_i \sim \text{Bernoulli}(\frac{1}{2})$											
	$\text{Indep } z_i \sim \text{Bernoulli}(\frac{1}{2})$				$z_i = x_i$				$z_i \sim \text{Bernoulli}(\frac{1}{2})I_{\{x_i > 0\}}$			
	$\tau^2 = 0.5$		$\tau^2 = 1$		$\tau^2 = 0.5$		$\tau^2 = 1$		$\tau^2 = 0.5$		$\tau^2 = 1$	
GLMM $\tilde{\beta}$	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_0$	$\tilde{\beta}_1$
GEE $\hat{\beta}$	1.13	1.01	1.28	1.00	1.00	1.24	1.00	1.49	1.00	1.13	1.04	1.24
GLMM $\hat{\beta}$	1.00	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
True $\tilde{\beta} - \beta$	0.13	0	0.28	0	0	0.25	0	0.50				
Ave. $\hat{\beta} - \tilde{\beta}$	0.13	-0.01	0.28	-0.01	0.00	0.25	0.00	0.49	0.00	0.13	0.04	0.23

Fonte: Zhang et al. (2012, p. 2075).

Com relação à situação (b):

- Em $\tau^2 = 0,5$, o parâmetro $\tilde{\beta}_0$ teve o mesmo valor estimado de β_0 , enquanto que $\tilde{\beta}_1$ expressou uma diferença (da ordem de 0,24) com relação ao valor estimado de β_1 , conforme esperado.
- Em $\tau^2 = 1$, o parâmetro $\tilde{\beta}_0$ manteve o mesmo valor estimado de β_0 , enquanto que, em $\tilde{\beta}_1$, a diferença aumentou (da ordem de 0,49) com relação ao valor estimado de β_1 .

Na situação (c), as diferenças estimadas entre $\tilde{\beta}$ e β sugerem, segundo os autores, que $\tilde{\beta}$ e β são diferentes em ambos os modelos. Mas, ainda assim, é possível notar uma maior similaridade da situação (c) com aquela observada em (b).

Além disso, os autores notaram que as verdadeiras diferenças esperadas para cada um dos parâmetros, $\text{True } \tilde{\beta} - \beta$, coincidiram praticamente em todos os casos com os valores estimados, Ave. $\hat{\beta} - \tilde{\beta}$.

Em seguida, Zhang et al. (2012) realizaram uma segunda rodada de simulações no intuito de analisar, em maiores detalhes, a relação entre os modelos sujeitos às condições (b) e (c).

Uma vez que $\tilde{\beta}_0$, em ambos os modelos, ficaram todos próximos de 1,00, os autores decidiram testar $H_0: \beta_0 = 1$, em ambas as classes de modelos (GLMM e GEE).

Em contrapartida, o teste a ser realizado em $\tilde{\beta}_1$ será dado por $H_0: \beta_1 = 0$. Isso porque os autores desejam avaliar o efeito $\frac{1}{2}\tau^2$, ora associado com o parâmetro $\tilde{\beta}_1$, conforme (6.71).

Em ambas as situações, foram realizadas simulações com $n = 100$ e $n = 500$, usando as estatísticas do TRV e do teste escore, inclusive com suas versões ajustadas para pequenas

amostras pela estatística T^2 de Hotelling, tanto para as estimativas dos parâmetros quanto para as estimativas das probabilidades do erro tipo I. Esses valores estão apresentados na Tabela 6.12.

É importante notar que, nessa segunda rodada de simulação, os autores estabelecem $\tilde{\beta}_1 = 0$, mas, que, apesar disso, em simulação, as estimativas desse parâmetro são permitidas. Nesse caso, foram realizadas 1.000 replicações.

Tabela 6.12 – Estimativas de parâmetros e de erros tipo I sob condições específicas

Table 2. Comparisons of parameter estimates and type I errors between the GEE and GLMM models based on data simulated from the GLMM model under Cases (b) and (c), with parameter estimates averaged over 1000 MC replications and type I errors computed based on LRT for the GLMM, score test (score) for GEE, and Hotelling's adjustment (Hot. adj.) for both models.

GLMM $\tilde{\beta}$		Case (b)				Case (c)			
		$\tau^2 = 0.5$		$\tau^2 = 1$		$\tau^2 = 0.5$		$\tau^2 = 1$	
		$\tilde{\beta}_0 = 1$	$\tilde{\beta}_1 = 0$	$\tilde{\beta}_0 = 1$	$\tilde{\beta}_1 = 0$	$\tilde{\beta}_0 = 1$	$\tilde{\beta}_1 = 0$	$\tilde{\beta}_0 = 1$	$\tilde{\beta}_1 = 0$
<i>Estimates from GEE and GLMM</i>									
$n = 100$	GEE $\hat{\beta}$	1.037	0.206	0.999	0.490	1.045	0.085	0.998	0.287
	GLMM $\hat{\beta}$	0.996	0.001	0.993	0.004	0.992	0.004	0.991	0.005
$n = 500$	GEE $\hat{\beta}$	0.999	0.248	0.998	0.497	0.999	0.131	1.000	0.277
	GLMM $\hat{\beta}$	0.997	0.003	0.997	0.002	0.998	-0.005	0.996	0.001
<i>Type I error for testing $H_0 : \tilde{\beta}_1 = 0$ (GLMM)/$\beta_1 = 0$ (GEE)</i>									
$n = 100$	LRT/score	0.066/0.486		0.052/0.747		0.046/0.194		0.032/0.387	
	Hot. adj.	0.065/0.479		0.049/0.740		0.043/0.190		0.030/0.378	
$n = 500$	LRT/score	0.055/0.983		0.057/1.000		0.055/0.741		0.041/0.982	
	Hot. adj.	0.054/0.983		0.056/1.000		0.053/0.740		0.041/0.981	

Fonte: Zhang et al. (2012, p. 2076).

Com relação às estimativas:

- Os valores de $\tilde{\beta}_0$ ficaram, em geral, próximos de 1, sendo que, em $n = 500$, essas estimativas ficaram todas bem mais próximas de 1.
- Com relação às estimativas de $\tilde{\beta}_1$, aquelas referentes ao modelo GLMM ficaram próximas de zero, mas, de outro modo, as estimativas do modelo GEE ficaram todas relativamente distantes de zero. Esse resultado condiz com o efeito $\frac{1}{2}\tau^2$ detectado no modelo marginal, inexistente no modelo original GLMM.

Com relação a esse último item, Zhang et al. (2012) sugerem que, apesar de desconhecido, o viés na situação (c) também deve ser elevado.

No segundo bloco da Tabela 6.12, os autores mediram as taxas das probabilidades de erro tipo I, especificamente para o teste de hipóteses $H_0: \beta_1 = 0$.

Tanto para GEE quanto para GLMM, Zhang et al. (2012) obtiveram as taxas das probabilidades de erro tipo I. Assim, para um teste (de hipóteses) de escore, por exemplo, a taxa das probabilidades de erro tipo I foi estimada por meio da expressão

$$\hat{\alpha} = \frac{1}{1000} \sum_{m=1}^{1000} I_{\{T_s^{(m)} \geq q_{1,0.95}\}}, \quad (6.72)$$

com $T_s^{(m)}$ representando a estatística do teste escore da m -ésima replicação, e $q_{1,0.95}$ denotando o 95º percentil da distribuição χ_1^2 . Assim, a função indicadora garante que $\hat{\alpha}$ represente a razão entre o número de estimativas de probabilidade de erro tipo I que ficaram acima do 95º percentil da distribuição χ_1^2 e o número total de estimativas. A versão adaptada da estatística T^2 de Hotelling para a obtenção de $\hat{\alpha}$ também foi utilizada.

Com relação às estimativas das probabilidades de erro tipo I, temos que:

- Para o parâmetro $\tilde{\beta}_1 = 0$, para os modelos GLMM e situações (b) e (c), as taxas das estimativas de $\hat{\alpha}$ ficaram próximas ao nível pré-estabelecido pelos autores, isto é, 0,05.
- Para o parâmetro $\tilde{\beta}_1 = 0$, para os modelos GEE e situações (b) e (c), as taxas das estimativas de $\hat{\alpha}$ permaneceram consideravelmente elevadas, em relação ao nível pré-estabelecido de 0,05.

Com base nesse segundo item, Zhang et al. (2012) sugerem que, no caso dos modelos GEE, as taxas de probabilidades de erros tipo I encontram-se fortemente viesadas, tanto em (b) quanto em (c), apesar do ajuste induzido pela estatística T^2 de Hotelling.

Ademais, os autores apresentaram mais uma tabela de resultados, desta vez comparando os métodos de estimação (MLE e SVE), tanto para as estimativas dos parâmetros quanto para as estimativas das probabilidades de erros tipo I, especificamente para o caso (a). Os dados foram gerados a partir de um modelo GLMM, com valores previamente fixados em $\beta_0 = 1,13$ e $\beta_1 = 1$, para $\tau^2 = 0,5$ e $\beta_0 = 1,28$ e $\beta_1 = 1$, para $\tau^2 = 1$, de acordo com valores estimados na Tabela 6.11. Os valores são apresentados na Tabela 6.13.

Com relação aos resultados da Tabela 6.13, temos que:

- Às estimativas dos parâmetros, os resultados ficaram dentro do esperado pelos autores, isto é, independentemente das diferenças dentre os métodos de estimação, as estimativas fornecidas foram as mesmas.
- Contudo, as taxas das probabilidades de erro tipo I, dentre os métodos de estimação, variaram consideravelmente.

Os autores apontam que o motivo das taxas das probabilidades de erro tipo I terem se mostrado consideravelmente infladas, sob o método de estimação de máxima verossimilhança,

é devido ao fenômeno da sobredispersão, mas que, quando o modelo é estimado via SVE, as taxas das probabilidades de erro tipo I permanecem próximas aos valores nominais de 0,05.

Tabela 6.13 – Comparação entre os métodos de estimação MLE e SVE

Table 3. Comparisons of parameter estimates and type I error rates from fitting Poisson loglinear model to overdispersed count data generated under the GLMM in Case (a) using MLE and GEE with SVE.

	$\tau^2 = 0.5$			$\tau^2 = 1$	
Sample size	True β	$\beta_0 = 1.13$	$\beta_1 = 1$	$\beta_0 = 1.28$	$\beta_1 = 1$
<i>Estimates of β from MLE and SVE</i>					
$n = 100$	SVE	1.131	0.999	1.273	0.995
	MLE	1.131	0.999	1.273	0.995
$n = 500$	SVE	1.129	1.002	1.276	1.001
	MLE	1.129	1.002	1.276	1.001
<i>Type I error for testing $H_0 : \beta_1 = 1$ from MLE/SVE</i>					
$n = 100$	LRT/score	0.229/0.055		0.434/0.069	
$n = 500$	LRT/score	0.277/0.061		0.490/0.056	

Fonte: Zhang et al. (2012, p. 2076).

6.4.2 Análise de Dados Reais

Zhang et al. (2012) realizaram, ainda, um estudo de ensaios clínicos sobre dependência de álcool com 1.383 indivíduos foi conduzido entre 2000 e 2004 pelo COMBINE Study Research Group. Duas variáveis respostas de interesse: i) número de dias sem beber e ii) número de dias com consumo elevado, durante o último mês.

Os dados foram coletados em um momento inicial e em três visitas durante o tratamento, $t = 1, 2, 3$, representando 8, 16 e 26 semanas, respectivamente, após o momento inicial. O modelo (de intercepto aleatório) levará em conta os dados dessas três visitas, de modo que

$$y_{it}|x_{it}, b_i \sim \text{Poisson}(\tilde{\mu}_{it}), \quad (6.73)$$

$$\log(\tilde{\mu}_{it}) = x_{i1t}\tilde{\beta}_0 + x_{i2t}\tilde{\beta}_1 + x_{i3t}\tilde{\beta}_2 + z_i b_i, \quad (6.74)$$

$$b_i \sim N(0, \tau^2), \quad (6.75)$$

$$x_{i1t} = 1 \quad \text{e} \quad x_{ikt} = \begin{cases} 1, & \text{se } t = k \\ 0, & \text{c. c.} \end{cases} \quad (6.76)$$

para $k = 2, 3$, $z_i = 1$, $1 \leq i \leq n$ e $1 \leq t \leq 3$.

Uma vez que x_{i1t} e z_i são constantes, também são independentes, contexto semelhante à situação (a) dos dados simulados, de forma que, podemos esperar que $\tilde{\beta}_0$ se diferencie de sua

contrapartida no modelo GEE (β_0) em $\gamma_0 = \ln \left(\mathbb{E} \left(\exp \left\{ \frac{1}{2} z_i \tau^2 \right\} \right) \right)$, enquanto que, para $\tilde{\beta}_1$ e $\tilde{\beta}_2$, não seriam esperadas diferenças com relação a β_1 e β_2 no GEE.

Os resultados obtidos pelos autores são apresentados na Tabela 6.14.

Conforme era esperado, as estimativas de $\tilde{\beta}_0$ e de β_0 diferiram significativamente (nos modelos de ambas as variáveis respostas), embora as diferenças tenham sido nulas em todas as demais estimativas ($\tilde{\beta}_1$, β_1 e $\tilde{\beta}_2$ e β_2).

Tabela 6.14 – Estimativas dos modelos GEE e GLMM de dados de alcoolismo

Comparison of estimates (standard errors $\times 10^{-2}$) between GEE and GLMM			
Models fit	Visit 1 (β_0 or $\tilde{\beta}_0$)	Visit 2 (β_1 or $\tilde{\beta}_1$)	Visit 3 (β_2 or $\tilde{\beta}_2$)
<i>Days of any drinking</i>			
GEE (β)	1.908(3.6)	0.144(2.2)	0.144(2.8)
GLMM ($\tilde{\beta}$)	0.694(6.2)	0.144(1.5)	0.144(1.5)
<i>Days of heavy drinking</i>			
GEE (β)	1.307(4.7)	0.224(3.3)	0.216(4.1)
GLMM ($\tilde{\beta}$)	-0.30(7.3)	0.224(1.9)	0.216(1.9)

Fonte: Zhang et al. (2012, p. 2077).

As estimativas dos erros padrão variaram significativamente entre os modelos GEE e GLMM em ambas as variáveis respostas, em um padrão sistemático, a saber, maior variabilidade, na visita 1, sempre para GLMM, e, nas visitas 2 e 3, a variabilidade é sempre maior nos modelos GEE, refletindo, dessa maneira, a sobredispersão.

De acordo, ainda, com Zhang et al. (2012), pelo fato de que os parâmetros $\tilde{\beta}_1$ e $\tilde{\beta}_2$ (GLMM) e β_1 e β_2 (GEE) descrevem mudanças ao longo do tempo a partir de $\tilde{\beta}_0$ (GLMM) e ou de β_0 (GEE), diferenças expressivas entre $\tilde{\beta}_0$ e β_0 produzem impacto não somente entre as estimativas iniciais (em $t = 1$), mas, também, ao longo do tempo ($t = 2, 3$).

Para o modelo cuja variável resposta é o número de dias sem beber, as estimativas do modelo GEE sugerem 6,7, 7,7 e 8,9 dias (em média) nas três visitas ($e^{1,908}$, $e^{1,908+0,144}$, $e^{1,908+0,144+0,144}$). Já no modelo GLMM, essas estimativas sugerem os valores 2, 2,3 e 2,7, os quais são bem diferentes quando comparados com as estimativas do modelo GEE.

Zhang et al. (2012) ressaltam, ainda, que as estimativas do modelo GEE correspondem, de fato, às verdadeiras médias em cada visita, de forma que o mesmo já não pode ser dito sobre as estimativas do modelo GLMM. Em razão disso, os autores chamam atenção para a dificuldade de uma conexão direta entre os dados observados e as estimativas do modelo GLMM.

6.5 Modelos PA e SS – Algumas relações gerais

Nesta seção, serão apresentadas algumas relações de equivalência entre modelos de regressão marginal e condicional analisadas por Ritz e Spiegelman (2004). Essas relações serão apresentadas tendo como base as funções de ligação mais comumente utilizadas, a saber, *identidade*, *log*, *logit* e *probit*.

De acordo com Ritz e Spiegelman (2004), a ênfase consistia em mostrar casos comuns sobre os quais parâmetros dos modelos seriam iguais ou de mesma interpretação.

Ritz e Spiegelman (2004) supõem, nesse artigo, independência entre as covariáveis e efeitos aleatórios, situação tipicamente encontrada na literatura.

De forma a caracterizar a relação geral entre os modelos (PA ou SS), Ritz e Spiegelman (2004) descrevem uma função da média condicional de Y_{ij} , com $j = 1, \dots, n_i$, sob a forma

$$g(\mathbb{E}[Y_{ij}|b_i]) = K[\mathbf{z}_i' \mathbf{b}_i, \mu(\mathbf{X}_i, \boldsymbol{\beta})], \quad (6.77)$$

a qual é descrita por uma função, K , de $\mathbf{z}_i' \mathbf{b}_i$ e de $\mu(\mathbf{X}_i, \boldsymbol{\beta})$.

A seguir, serão descritos quatro resultados apontados em Ritz e Spiegelman (2004), acerca das condições de equivalência entre modelos condicionais e marginais.

a) Regressão marginal é equivalente a regressão condicional quando os efeitos aleatórios são aditivos.

Para esse resultado, Ritz e Spiegelman (2004) citam, como exemplo, um modelo abordado em Neuhaus et al. (1991) e em Neuhaus e Jewell (1993) da classe LMM, isto é, um modelo com resposta normalmente distribuída, conforme (6.78).

$$Y_{ij}|b_i = \beta_0 + X_{ij}\beta_1 + b_i + \varepsilon_{ij}. \quad (6.78)$$

Nesse caso, os ε_{ij} 's são independentes com distribuição $N(0, \sigma_\varepsilon^2)$ e os b_i 's também são independentes, mas, distribuídos segundo $N(0, \tau^2)$.

Ao relacionar o modelo (6.78) com a relação (6.77), temos que $\mu(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$, $g(\mathbb{E}[Y_{ij}|b_i]) = K(\mathbf{b}, \mathbf{X}\boldsymbol{\beta}) = \mathbf{b} + \mathbf{X}\boldsymbol{\beta}$, sendo que $\mu(\mathbf{X}, \boldsymbol{\beta})$ é linear tanto nos parâmetros quanto na covariável e $\mathbf{b}$ é aditivo em relação a μ .

Na realidade, Ritz e Spiegelman (2004) chamam atenção para o fato de que o modelo (6.78) constitui um caso particular de um resultado mais geral, a saber, de que se um modelo para a média condicional de $\mathbf{Y}$, sob a forma (6.77), for aditivo em $\mathbf{b}$ e em $\mu(\mathbf{X}, \boldsymbol{\beta})$, segue, então, que, para qualquer distribuição de $\mathbf{Y}$ (cuja média exista), a média marginal de $\mathbf{Y}$ será dada pela fórmula iterativa

$$\begin{aligned}\mathbb{E}[\mathbf{Y}] &= \mathbb{E}_b[\mathbb{E}[\mathbf{Y}|\mathbf{b}]] \\ &= \mathbb{E}_b[\mu + \mathbf{b}] \\ &= \mu + \mathbb{E}[\mathbf{b}],\end{aligned}\tag{6.79}$$

com $\mu = \mu(\mathbf{X}, \boldsymbol{\beta})$ e independência entre os efeitos aleatórios e $\mathbf{X}$. Além disso, Ritz e Spiegelman (2004) observam que (6.79) é válido para qualquer função $\mu = h(\mathbf{X}, \boldsymbol{\beta})$, sendo a função h não necessariamente restrita à forma linear e para qualquer distribuição F_b de $\mathbf{b}$ cuja média exista.

Assim, os autores estabelecem que, para qualquer distribuição condicional $\mathbf{Y}|\mathbf{b}$, para qualquer efeito aleatório $\mathbf{b}$ com média zero e independente de $\mathbf{X}$, se a média condicional (6.77) for aditiva em μ e em $\mathbf{b}$, segue que, em (6.79), teremos $\mathbb{E}[\mathbf{Y}] = \mu$.

Nesse caso, a variância marginal seria obtida por

$$\begin{aligned}\text{Var}[\mathbf{Y}] &= \mathbb{E}_b[\text{Var}[\mathbf{Y}|\mathbf{b}]] + \text{Var}_b[\mathbb{E}[\mathbf{Y}|\mathbf{b}]] \\ &= \mathbb{E}_b[\text{Var}[\mathbf{Y}|\mathbf{b}]] + \text{Var}[\mathbf{b}].\end{aligned}\tag{6.80}$$

Segundo Ritz e Spiegelman (2004), o aspecto fundamental para a equivalência entre os modelos condicionais e marginais reside na aditividade entre $\mathbf{b}$ e $\mu(\mathbf{X}, \boldsymbol{\beta})$, conforme as relações que resultaram em (6.79), isto é, quando a média condicional $\mathbb{E}[\mathbf{Y}|\mathbf{b}]$ for aditiva em μ e em $\mathbf{b}$, então, temos que $\mathbb{E}[\mathbf{Y}|\mathbf{b}] = \mathbb{E}_b[\mu + \mathbf{b}] = \mu + \mathbb{E}[\mathbf{b}]$.

Os autores afirmam, ainda, que a hipótese de normalidade para os efeitos aleatórios não é necessária para a validade dessa regra, mesmo para formas não lineares de h .

A título de ilustração, os autores citam um exemplo em Diggle, Liang e Zeger (1994) em que Y dado b admite distribuição de Poisson com média b , tendo os efeitos aleatórios, b , distribuição Gama com média γ e variância $\phi\gamma^2$. A distribuição marginal de Y , ainda que com distribuição Binomial Negativa e variância $\gamma + \phi\gamma^2$, também admitirá, de modo equivalente, a mesma média, γ , ou seja, verifica-se que $\mathbb{E}[Y] = \mu$.

Um segundo exemplo ilustrativo contido em Ritz e Spiegelman (2004) diz respeito à distribuição condicional de Y , dado b também admitir distribuição de Poisson, mas com média $\mu + b$, com $b \sim N(0, \sigma^2)$, e função de ligação *identidade* dada por $\mu = \beta_0 + X\boldsymbol{\beta}$. Nesse caso, a média marginal de Y também seria igual a μ .

b) Quando a média condicional é aditiva em um efeito aleatório com função de ligação log, a média marginal é igual à média condicional mais uma constante.

Ainda que o resultado anterior (a) não seja restrito apenas para o caso de funções de ligação lineares, Ritz e Spiegelman (2004) abordam um resultado especificamente para a função de ligação *log*. Nesse caso, os autores assumem que $\log(\mathbb{E}[Y|X]) = \mu(X, \beta) + b$, de modo que a média marginal para Y seja dada por

$$\mathbb{E}[Y] = \int e^{\mu(X, \beta) + b} dF_b(b). \quad (6.81)$$

Na medida em que a integral de $e^b dF_b(b)$ exista e seja finita, a média marginal será igual à média condicional mais uma constante que depende dos parâmetros da distribuição dos efeitos aleatórios.

Essa constante será incorporada no intercepto da média marginal, sempre que $\mu(X, \beta)$ for uma função linear. Quaisquer que sejam a distribuição de F_b e a distribuição condicional de Y , dados X e b , esse resultado será válido, desde que a média condicional seja aditiva na escala (da função de ligação) *log*.

A título de ilustração, Ritz e Spiegelman (2004) citaram o modelo de Breslow e Clayton (1993) tal que Y_i é condicionalmente independente, dado o vetor de efeitos aleatórios $\mathbf{b}$, de dimensão $q \times 1$, com distribuição normal multivariada de média $\mathbf{0}$ e matriz de covariância $\mathbf{D}$.

Esse resultado (b) fica mais evidente à medida que são introduzidos vetores $\mathbf{z}_i$'s sobre os efeitos aleatórios $\mathbf{b}$. Nesse caso, a média condicional seria descrita por $\mathbb{E}[Y_i|\mathbf{b}] = \mu_i^{(b)}$, sendo que $\mu_i^{(b)} = h(\mathbf{X}_i'\beta + \mathbf{z}_i'\mathbf{b})$, onde h é a inversa da função de ligação g .

Com uma função de ligação *log*, teríamos $\mathbb{E}[Y_i] = \exp\{\mathbf{X}_i'\beta + \mathbf{z}_i'\mathbf{b}/2\}$, de forma que é possível notar que a variância do efeito aleatório acabou por introduzir uma compensação (*offset*) na média marginal.

Ritz e Spiegelman (2004) destacam também que nenhuma suposição sobre a distribuição de probabilidades dos efeitos aleatórios foi necessária, condição essa válida tanto para a função de ligação *identidade* quanto para a função de ligação *log*. A única exigência é que tanto os efeitos fixos quanto os efeitos aleatórios sejam aditivos e que o valor esperado de $h(\mathbf{z}_i'\mathbf{b})$ exista e seja finito.

c) Relação entre média marginal e média condicional funciona quando $\mathbf{b}$ é aditivo em $\mu(X, \beta)$, com função de ligação *probit*.

O exemplo citado por Ritz e Spiegelman (2004) baseia-se no próprio exemplo abordado em Heagerty e Zeger (2000), artigo que, conforme mencionado anteriormente, desvinculou o método de estimação de escolha de modelos (PA e SS).

Segundo Ritz e Spiegelman (2004), o modelo abordado em Heagerty e Zeger (2000) é dado por $\text{probit}(\mathbb{E}[Y_i|\mathbf{b}]) = \mu(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{b}$, também, dentro das conformidades de (6.77), tendo que a função de ligação g é a função *probit*.

No caso, Heagerty e Zeger (2000, p. 7) notaram que, supondo efeitos aleatórios normalmente distribuídos e função de ligação *probit*, o preditor linear do modelo marginal seria um re-escalamento do preditor linear do modelo condicional ou, conforme Ritz e Spiegelman (2004), os parâmetros do modelo marginal seriam atenuados por um fator que dependeria de $\mathbf{X}$.

Em termos mais específicos, teríamos $\mathbb{E}[Y] = h[\mu(\mathbf{X}, \boldsymbol{\beta})/c]$, sendo $c = \sqrt{1 + \text{Var}(\mathbf{b})}$ e $h = g^{-1}$, sob a forma da distribuição normal acumulada, Φ .

Heagerty e Zeger (2000, p. 7) chamam atenção, ainda, para o fato de que, nesse contexto específico de função de ligação *probit*, se a variância dos efeitos aleatórios é independente das covariáveis, então, as estruturas dos modelos marginais e condicionais serão as mesmas. Do contrário, assumirão formas distintas.

d) As condições para a equivalência entre as médias marginais e condicionais funcionam quando $\mathbf{b}$ é aditivo em $\mu(\mathbf{X}, \boldsymbol{\beta})$, na escala *logit*.

Ritz e Spiegelman (2004) citam o exemplo de Neuhaus et al. (1991) já analisado neste capítulo, qual seja o exemplo do modelo logístico de efeitos mistos com função de ligação *logit* descrito na expressão (6.15).

Conforme descrito acerca da expressão (6.15), apesar da integral representar um modelo marginal, esse não admitiria, a rigor, a forma logística, o que impediria, em certa medida, a realização de comparação entre os parâmetros dos modelos PA e SS.

Ritz e Spiegelman (2004) observaram que os resultados obtidos pela metodologia utilizada em Neuhaus et al. (1991) também seriam válidos para um efeito aleatório adicional (em um coeficiente β_i com $i = 1, \dots, k$), não se restringindo somente ao intercepto.

Assim, para uma função de ligação *logit* e considerando efeito aleatório, $\mathbf{b}$, aditivo, a expressão (6.15) de Neuhaus et al. (1991) seria descrita sob a forma

$$\mathbb{E}[Y] = \int_{\mathbf{b}} (1 + e^{-(\mu + \mathbf{b})})^{-1} dF_{\mathbf{b}}(\mathbf{b}), \quad (6.82)$$

ainda que Ritz e Spiegelman (2004) reconheçam que, para efeitos aleatórios com distribuição normal, uma forma fechada para a expressão (6.82) não seja de fácil obtenção.

Ritz e Spiegelman (2004) apontam uma solução apresentada em Crouch e Spiegelman (1990), a saber, um algoritmo numérico capaz de efetuar o cálculo dessa integral, com grau pré-especificado de acurácia. Dessa forma, a solução dessa integral permitiria estimar o modelo GEE e, conseqüentemente, a devida comparação entre os parâmetros PA e SS.

Segundo, ainda, Ritz e Spiegelman (2004), os autores Knox, Spiegelman, Skinner et al. (2000) teriam utilizado esse mesmo algoritmo em um contexto no qual a função de ligação *logit* também teria sido empregada, com atenuação de β . A expressão para a média marginal assumiria a forma

$$\mathbb{E}[Y] \approx \frac{\exp\{a_i \mathbf{X}'_i \beta\}}{1 + \exp\{a_i \mathbf{X}'_i \beta\}}, \quad (6.83)$$

tal que $a_i = |\mathbf{I} + c^2 \mathbf{D} \mathbf{z}_i \mathbf{z}'_i|^{-1/2}$, $c = 16\sqrt{3}/(15\pi)$ e a_i entre 0 e 1.

6.6 Apêndice

- **Demonstração da existência de sobredispersão em um modelo marginal obtido a partir de um modelo condicional com resposta de Poisson e função de ligação log.**

Pela lei da variância total, a variância marginal de $\mathbf{y}$ pode ser obtida por

$$\begin{aligned} \text{Var}(y_i) &= \text{Var}(\mathbb{E}[y_i | \mathbf{b}]) + \mathbb{E}[\text{Var}(y_i | \mathbf{b})] \\ &= \text{Var}(\mu_i) + \mathbb{E}[g(\mu_i) \cdot \phi], \end{aligned}$$

uma vez que $\mathbb{E}[y_i | \mathbf{b}] = \mu_i$ e $\text{Var}(y_i | \mathbf{b}) = g(\mu_i) \cdot \phi$, onde $g(\mu_i)$ é a *função variância*.

Se for feita a suposição de função de ligação *log*, será possível, por meio da inversa dessa função e para $\mathbf{y}$, dado $\mathbf{b}$, independentes, assumir que, em um contexto de distribuição de Poisson, $\mu_i = g(\mu_i) \phi$. Dessa maneira,

$$\text{Var}(y_i) = \text{Var}(\mu_i) + \mathbb{E}[\mu_i].$$

Uma vez que $\mu_i = \exp\{x'_i \beta + z'_i \mathbf{b}\}$, temos que

$$\text{Var}(y_i) = \text{Var}(\exp\{x'_i \beta + z'_i \mathbf{b}\}) + \mathbb{E}[\exp\{x'_i \beta + z'_i \mathbf{b}\}].$$

Lembrando que $\text{Var}(X) = \mathbb{E}[X^2] - [\mathbb{E}(X)]^2$, temos que

$$Var(y_i) = \mathbb{E}[(\exp\{2(x'_i\boldsymbol{\beta} + z'_i\mathbf{b})\})] - [\mathbb{E}(\exp\{x'_i\boldsymbol{\beta} + z'_i\mathbf{b}\})]^2 + \mathbb{E}[\exp\{x'_i\boldsymbol{\beta} + z'_i\mathbf{b}\}],$$

lembrando que $\mathbb{E}[e^X]^2 = \mathbb{E}[e^{2X}]$.

Recordando do resultado (6.28), temos que

$$\begin{aligned} Var(y_i) &= \exp\{2x'_i\boldsymbol{\beta}\}M_b(2z_i) - [\exp\{x'_i\boldsymbol{\beta}\}M_b(z_i)]^2 + \exp\{x'_i\boldsymbol{\beta}\}M_b(z_i) \\ &= \exp\{2x'_i\boldsymbol{\beta}\} \left[M_b(2z_i) - (M_b(z_i))^2 + \exp\{-x'_i\boldsymbol{\beta}\}M_b(z_i) \right]. \end{aligned}$$

Assumindo que $b_i \sim N(0, \sigma_b^2)$, e que em cada uma das linhas de Z tenha uma única entrada igual a 1, com todos os demais elementos da mesma linha iguais a zero (correspondente ao intercepto aleatório), então,

$$\begin{aligned} Var(y_i) &= \exp\{2x'_i\boldsymbol{\beta}\}[\exp\{2\sigma_b^2\} - (\exp\{\sigma_b^2/2\})^2 + \exp\{-x'_i\boldsymbol{\beta}\}\exp\{\sigma_b^2/2\}] \\ &= \exp\{2x'_i\boldsymbol{\beta} + \sigma_b^2/2\}[\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\} + \exp\{-x'_i\boldsymbol{\beta}\}] \\ &= \exp\{x'_i\boldsymbol{\beta} + \sigma_b^2/2\}[\exp\{x'_i\boldsymbol{\beta}\}(\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\} + \exp\{-x'_i\boldsymbol{\beta}\})] \\ &= \exp\{x'_i\boldsymbol{\beta} + \sigma_b^2/2\}[\exp\{x'_i\boldsymbol{\beta}\}(\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\}) + 1] \\ &= \mathbb{E}(y_i)[\exp\{x'_i\boldsymbol{\beta}\}(\exp\{3\sigma_b^2/2\} - \exp\{\sigma_b^2/2\}) + 1]. \end{aligned}$$

Uma vez que a expressão entre colchetes é notadamente maior que 1, temos que $Var(y_i) > E(y_i)$, de modo que, nesse caso, a distribuição marginal apresentará sempre sobredispersão. Isso significa, também, que a distribuição marginal não possui distribuição de Poisson, apesar da distribuição condicional admitir distribuição de Poisson.

7 SIMULAÇÕES E ANÁLISES COMPARATIVAS ENTRE OS MODELOS PA E SS

7.1 Introdução

Conforme analisado no capítulo anterior, Zeger et al. (1988) sugeriram duas estratégias para lidar com a análise de dados longitudinais: i) Modelos SS (*Subject-Specific*) e ii) Modelos PA (*Population-Averaged*), sendo que as abordagens para essas estratégias seriam os modelos de média condicional e de média marginal, respectivamente.

Foi analisado, também, que, de um modo geral, a integral da média condicional, na função densidade de probabilidade dos efeitos aleatórios, resulta no modelo da média marginal.

Além disso, Zeger et al. (1988) verificaram algumas relações matemáticas aproximadas entre os parâmetros dos modelos PA e SS, dispostas na Tabela 6.1, com exemplo aplicado em modelo de regressão logística. Essas relações seriam válidas para as funções de ligação *identidade*, *log*, *logit* e *probit*.

Entretanto, Young et al. (2007) estenderam essa análise para o caso da regressão de Poisson, com função de ligação *log*.

Neste capítulo, procuraremos, por meio uma simulação, realizar uma aplicação desses conceitos no intuito de comparar o grau de aproximação dos parâmetros dos modelos PA e SS, tanto para a regressão logística quanto para a regressão de Poisson.

7.2 As estratégias PA e SS

O modelo SS descrito em Zeger et al. (1988) considerava uma variável resposta y_{it} e um vetor de covariáveis $\mathbf{x}_{it}$ fixas e de ordem $p \times 1$, respectivamente, no tempo t e para a i -ésima UE, tal que $t = 1, \dots, n_i$ e $i = 1, \dots, N$.

Além disso, foi considerado um vetor $\mathbf{z}_{it}$, de dimensão $q \times 1$ de covariáveis (possivelmente um subconjunto de $\mathbf{x}_{it}$), associado a um vetor (também de dimensão $q \times 1$) de coeficientes dos efeitos aleatórios, $\mathbf{b}_i$, de forma que, para o modelo de média condicional da i -ésima UE,

$$\mu_{it}^{ss} = \mathbb{E}(y_{it}|\mathbf{b}_i), \quad (7.1)$$

com uma função de ligação capaz de relacionar

$$h(\mu_{it}^{ss}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i, \quad (7.2)$$

e função de variância $\text{Var}(y_{it}|\mathbf{b}_i) = g(\mu_{it}^{ss})\phi$.

Por meio da função inversa de $h(\cdot)$, seria possível escrever $\mu_{it}^{SS} = h^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i)$, de forma que a média marginal μ_{it}^{PA} , a partir da média condicional μ_{it}^{SS} , seria obtida pela integral

$$\mu_{it}^{PA} = \mathbb{E}(y_{it}) = \mathbb{E}[\mathbb{E}(y_{it}|\mathbf{b}_i)] = \int h^{-1}(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i) dF(\mathbf{b}_i), \quad (7.3)$$

considerando, ainda, os efeitos aleatórios com distribuição normal.

7.2.1 Estratégia PA e SS: o Modelo de Regressão de Poisson

Nesta seção, será abordada, de forma resumida e a título de ilustração, uma relação simples para a relação entre os modelos PA e SS para a regressão de Poisson e com função de ligação *log*, baseada em Young et al. (2007). Esse resultado refere-se ao modelo Pós-teste.

Suponha um modelo GLMM com resposta de Poisson e efeitos aleatórios $b_i \sim N(0, \sigma_b^2)$. No caso de uma única covariável, o modelo para a média condicional seria dado por

$$\log(\mathbb{E}(y_{it}|\mathbf{b}_i)) = \log(\mu_{it}^{SS}) = \beta_0^{SS} + \beta_1^{SS}X_i + b_i, \quad (7.4)$$

sendo que, por meio da função inversa, teríamos $\mathbb{E}(y_{it}|\mathbf{b}_i) = \mu_{it}^{SS} = \exp\{\beta_0^{SS} + \beta_1^{SS}X_i + b_i\}$.

Por (7.3), sabemos que o modelo para a média marginal pode ser obtido por

$$\begin{aligned} \mu_{it}^{PA} &= \mathbb{E}[\mathbb{E}(y_{it}|\mathbf{b}_i)] \\ &= \mathbb{E}[\exp\{\beta_0^{SS} + \beta_1^{SS}X_i + b_i\}] \\ &= \exp\{\beta_0^{SS} + \beta_1^{SS}X\} \mathbb{E}[\exp\{b_i\}] \\ &= \exp\{\beta_0^{SS} + \beta_1^{SS}X\} \exp\{\sigma_b^2/2\}, \end{aligned} \quad (7.5)$$

de modo que

$$\log(\mu_{it}^{PA}) = (\beta_0^{SS} + \sigma_b^2/2) + \beta_1^{SS}X, \quad (7.6)$$

correspondendo ao modelo da média marginal com função de ligação *log* da Tabela 6.1, para o caso de uma única covariável X , sendo $\mathbf{z}_{it} = 1$ (modelo de intercepto aleatório).

Isso significa que, por um lado, $\beta_1 = \beta_1^{SS} = \beta_1^{PA}$ possuem todos a mesma interpretação, enquanto que, por outro lado, β_0^{SS} no modelo GLMM (7.4) é menor que o seu correspondente no modelo PA (7.6), em $\sigma_b^2/2$.

7.2.2 Estratégia PA e SS: o Modelo de Regressão Logística com Função de Ligação *logit*

Zeger et al. (1988) salientaram a inexistência de uma forma fechada no que se refere ao modelo de média marginal (PA) em um contexto de regressão logística com função de ligação *logit*. No entanto, ao utilizar uma aproximação (gaussiana acumulada) tal qual a descrita na

Tabela 6.1 por meio do resultado de Johnson e Kotz (1970), Zeger et al. (1988) obtiveram a expressão $\text{logit}(\mu_{it}) \approx a_i(\mathbf{D}) \cdot \mathbf{x}_{it}\boldsymbol{\beta}$, na qual

$$a_i(\mathbf{D}) = |c^2 \mathbf{D} \mathbf{z}_{it} \mathbf{z}_{it}' + \mathbf{I}|^{-q/2}, \quad (7.7)$$

sendo $c = \frac{16\sqrt{3}}{15\pi}$ e q a dimensão de $\mathbf{b}_i$.

Szmaragd, Clarke e Steele (2013, p. 159) mencionam uma regra para interpretar parâmetros PA a partir de parâmetros SS em um contexto de regressão logística, tanto para interceptos aleatórios quanto para coeficientes aleatórios. Essa regra seria dada pela relação aproximada

$$\beta^{PA} \approx \sqrt{\frac{3,29}{3,29 + \sigma_b^2}} \beta^{SS}, \quad (7.8)$$

na qual 3,29 seria a variância de uma distribuição logística padrão e σ_b^2 , a variância do efeito aleatório, de forma que, por meio da expressão (7.8), seria possível afirmar que não haveria diferença de interpretação entre os coeficientes PA e SS em caso de $\sigma_b^2 = 0$, ou seja, se não houvesse variabilidade no interior das UEs.

Szmaragd et al. (2013, pp. 160; 165) observam, ainda, que essa relação pode ser estendida para estruturas mais complexas de efeitos aleatórios, tais como modelos com interceptos e coeficientes aleatórios, citando, também, para isso, o próprio resultado do artigo de Zeger et al. (1988), a saber,

$$\beta^{PA} = \frac{\beta^{SS}}{\sqrt{1 + 0,346 \cdot \sigma_b^2}}, \quad (7.9)$$

o qual coincide com o resultado apresentado na Tabela 6.1 para a função de ligação *logit*, levando em consideração o fato de que $\left(\frac{16\sqrt{3}}{15\pi}\right)^2 \approx 0,346$.

Essa aproximação, descrita sob a forma (7.9), também aparece em Long, Loeber e Farrington (2009) e em Nakagawa, Johnson e Schielzeth (2018).

7.2.3 Relação entre os Modelos SS e PA: Regressão Logística com Função de Ligação *probit*

De acordo com os resultados de Zeger et al. (1988) contidos na Tabela 6.1, a expressão para a média marginal com função de ligação *probit* diferencia-se da expressão para a média marginal com função de ligação *logit* apenas pela presença (ou ausência) da constante c^2 . Logo,

a relação entre os parâmetros dos modelos PA e SS com função de ligação *probit* passa a ser dada por

$$\beta^{PA} = \frac{\beta^{SS}}{\sqrt{1 + \sigma_b^2}}. \quad (7.10)$$

Nakagawa et al. (2018) e Hedeker du Toit, Demirtas e Gibbons (2018) também utilizam o resultado de Zeger et al. (1988) sob a forma (7.10).

7.3 Simulações e comparações dos parâmetros PA e SS

Nesta seção, serão conduzidas atividades de: i) Simulação de dados longitudinais, ii) Estimação de modelos PA e SS e iii) Comparação dos parâmetros dos modelos PA e SS.

Para tal, será utilizado o software R, bem como os pacotes *simstudy*, *geepack*, *MASS* e *nlme*.

Por meio do pacote *simstudy*, é possível realizar a simulação de dados longitudinais envolvendo medidas repetidas. Esse tipo de exercício propicia aos pesquisadores, de um modo geral, uma atividade mais intensa (e independente) de investigação e análise de modelos longitudinais, uma vez que, ao pesquisador, é dado o poder de escolha das características tanto das variáveis resposta quanto das variáveis explicativas.

O pacote *geepack* foi escolhido para estimar modelos PA (GEE). Já os pacotes *MASS* e *nlme* foram utilizados para a estimação de modelos SS (GLMM). O pacote *MASS* possui um comando que permite estimar modelos GLMM pelo método PQL (comando *glmmPQL*). No caso do pacote *nlme* (juntamente com *MASS*), existe a vantagem de adicionar estruturas de correlações, bem como tornar possível um maior controle do próprio processo de estimação (modificar o otimizador, o número de iterações etc).

De um modo geral, a ideia de utilizar o pacote *simstudy* consiste em planejar e simular um experimento com intervenção. Para a formação do banco de dados, serão gerados 10.000 registros de id's, e um mecanismo de seleção balanceado (do próprio pacote *simstudy*) foi utilizado para designar a condição de cada um dos id's (tratamento ou controle). Em seguida, foram criados seis períodos longitudinais (variando de 0 a 5), implicando em um tamanho total de 60.000 linhas.

Os bancos de dados simulados são de dois tipos:

- 1) com variável resposta de Poisson;
- 2) com variável resposta binária.

Em ambos os casos, foi possível estabelecer correlação de 0,6 entre as medidas repetidas com estruturas “ar1” (autorregressivo de ordem 1) e “cs” (*compound symmetry*). Com relação às funções de ligação, no caso dos modelos PA e SS com resposta de Poisson, a função de ligação *log* será utilizada.

Já para os modelos com respostas binárias, duas funções de ligação serão utilizadas: a) *logit* e b) *probit*. Dessa maneira, serão três grupos de análise a serem conduzidos:

Tabela 7.1 – Grupos de Análise

Grupo	Dados simulados	Modelo	Classes PA e SS	Função de ligação
1	Poisson	Regressão de Poisson	GEE e GLMM	<i>log</i>
2	binários	Regressão Logística	GEE e GLMM	<i>logit</i>
3	binários	Regressão Logística	GEE e GLMM	<i>probit</i>

Fonte: O autor.

É importante destacar que, especificamente para o caso dos modelos PA, a estrutura de correlação “cs” (*compound symmetry*) foi utilizada para a simulação, enquanto que, para a estimação dos modelos (GEE), a estrutura “exchangeable” de correlação foi utilizada. Isso, porque, nos pacotes pesquisados no R para modelagem GEE, em geral, era oferecida a estrutura “exchangeable”. Contudo, é importante enfatizar que existe uma relação de equivalência entre essas estruturas de correlação. No caso da estrutura de correlação “cs”, a qual é dada por

$$\mathbf{R}_i = \begin{bmatrix} \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \dots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2 + \sigma_e^2 \end{bmatrix}, \quad (7.11)$$

com σ_b^2 e σ_e^2 representando as medidas de variância entre e intra-UEs, respectivamente, de forma que a soma $\sigma_b^2 + \sigma_e^2$ sintetiza a variabilidade total de uma UE.

Desse modo, ao considerar $\sigma^2 = \sigma_b^2 + \sigma_e^2$, é possível estabelecer uma estrutura de correlação denominada “exchangeable”, dada por

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}, \quad (7.12)$$

tal que $\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$.

Acrescentamos, ainda, que a ênfase das análises incidirá mais sobre a aproximação entre os parâmetros dos modelos das classes PA e SS.

7.3.1 Relação entre PA-SS via Regressão de Poisson com Função de Ligação *log*

Os dados analisados nos modelos Pré-teste/Pós-teste em Young et al. (2007) serviram de inspiração para essa simulação. Para a base de dados (xbase) desta seção, buscou-se, inicialmente, simular 10.000 id's, todos com distribuição Gama, de mesma média (igual a 10) e variância (igual a 0,5). O “sorteio” de 5.000 id's para tratamento e 5.000 id's para controle foi realizado por meio de um comando específico e estaria codificado (0 ou 1) na coluna “Trt”. Em seguida, uma fórmula foi criada para representar um efeito de intervenção (xbase + period * Trt) de forma que aqueles que estivessem no grupo de tratamento apresentariam, em razão da fórmula, valores crescentes e maiores ao longo do tempo.

O resultado dessa fórmula estaria em uma coluna denominada “lambda”, a qual serviria como parâmetro para simulação de uma nova variável aleatória denominada “NewPois”, com correlação igual a 0,6 (estrutura de correlação compound symmetry) e controlada por id's.

Com relação aos modelos SS (GLMM-cs e GLMM-ar1) a serem estimados, sua expressão assume a forma

$$\begin{aligned} \log(\mu_{ijt}) = & (\beta_0^{SS} + b_0) + \beta_1^{SS}(Treat) + (\beta_2^{SS} + b_2)(period) \\ & + \beta_3^{SS}(Treat \times period), \end{aligned} \quad (7.13)$$

com intercepto e coeficiente aleatórios dados, respectivamente, por b_0 e b_2 , conforme modelo descrito em Young et al. (2007).

Logo, a expressão para a média marginal, a qual servirá de referência para comparação entre os parâmetros dos modelos PA e SS, será dada por

$$\begin{aligned} \log(\mu_{it}) = & \left(\beta_0^{SS} + \frac{\sigma_0^2}{2} \right) + \beta_1^{SS}(Treat) + \left(\beta_2^{SS} + \frac{\sigma_2^2}{2} \right)(period) \\ & + \beta_3^{SS}(Treat \times period), \end{aligned} \quad (7.14)$$

de modo que β_0 (intercepto) e β_2 (coeficiente *period*) podem se comparados diretamente por meio de $\beta_0^{PA*} \approx \beta_0^{SS} + \frac{\sigma_0^2}{2}$ e por $\beta_2^{PA*} \approx \beta_2^{SS} + \frac{\sigma_2^2}{2}$.

As estimativas encontram-se disponíveis na Tabela 7.2. Com base na Tabela 7.2, é possível notar que, ao nível de $\alpha = 0,05$, o intercept e os coeficiente *Treat* e de interação

mostraram-se estatisticamente significantes em ambos os modelos (p-valor < 0,05). Entretanto, o coeficiente *period* mostrou-se significativo apenas no modelo GEE-exch (p-valor < 0,05). No modelo GEE-ar1, seu p-valor estimado foi de 0,0851.

As variâncias dos efeitos aleatórios (*intercept* e *period*) foram devidamente estimadas em ambos os modelos. Essas estimativas serão utilizadas para avaliar a aproximação entre os parâmetros dos modelos PA e SS, em ambas as estruturas de correlação.

Tabela 7.2 – Estimativas dos modelos SS (regressão de Poisson e função de ligação *log*)

Método	Parâmetro	Estimativa	SE	t-valor	p-valor
GLMM-cs	Intercepto	2,28888	0,0102349	223,634	0,0000
	factor(Treat)1	0,03652	0,0143146	2,551	0,0107
	period	-0,00164	0,0007010	-2,344	0,0191
	factor(Treat)1:period	0,08169	0,0010035	81,400	0,0000
	σ_0^2	6,9931e-09			
	σ_2^2	4,7993e-05			
GLMM-ar1	Intercepto	2,11070	0,0098783	213,671	0,0000
	factor(Treat)1	0,08705	0,0138450	6,288	0,0000
	period	-0,00228	0,0013221	-1,722	0,0851
	factor(Treat)1:period	0,08234	0,0017735	46,426	0,0000
	σ_0^2	0,34438			
	σ_2^2	3,1396e-10			

Fonte: O autor.

Para a estimação do modelo marginal, PA, com resposta de Poisson e estruturas de correlação “*exchangeable*” e “*ar1*” e função de ligação *log*, o modelo será dado por

$$\log(\mu_{ijt}) = \beta_0^{PA} + \beta_1^{PA}(Treat) + \beta_2^{PA}(period) + \beta_3^{PA}(Treat \times period) \quad (7.15)$$

O resumo das estimativas da classe de modelos PA (GEE-exch e GEE-ar1) encontra-se na Tabela 7.3.

Com base na Tabela 7.3, é possível notar que, ao nível de $\alpha = 0,05$, o coeficiente *Treat* mostrou-se estatisticamente significativo em ambos os modelos (p-valor < 0,05). No entanto, o coeficiente *period* mostrou-se significativo apenas no modelo GEE-exch (p-valor < 0,05). No modelo GEE-ar1, seu p-valor estimado foi de 0,068.

As correlações estimadas em ambos os modelos (0,922 e 0,95) ficaram acima do estabelecido na simulação (0,6). O parâmetro de dispersão, ϕ , também ficou elevado em ambos os modelos (5,37 e 5,42), sugerindo evidências de sobredispersão em ambos os modelos.

Tabela 7.3 – Estimativas dos modelos PA (regressão de Poisson e função de ligação $\log$)

Método	Parâmetro	Estimativa	SE	Wald	Pr(> W)
GEE-exch	Intercepto	2,28347	0,01075	45153,20	< 2e-16
	factor(Treat)1	0,03492	0,01487	5,51	0,019
	period	-0,00157	0,00069	5,19	0,023
	factor(Treat)1:period	0,08176	0,00112	5318,85	< 2e-16
	ϕ	5,37			
	α	0,922			
GEE-ar1	Intercepto	2,28268	0,01091	43745,55	< 2e-16
	factor(Treat)1	0,02373	0,01519	2,44	0,118
	period	-0,00225	0,00123	3,33	0,068
	factor(Treat)1:period	0,08225	0,00177	2171,39	< 2e-16
	ϕ	5,42			
	α	0,95			

Fonte: O autor.

Assim, por exemplo, no que diz respeito à aproximação do intercepto do modelos PA e SS (modelo GLMM-exch como referência), temos que $\beta_0^{SS} + \frac{\sigma_0^2}{2} = 2,28888 + \frac{6,9931e-09}{2} = 2,28888$, tendo ficado muito próximo a $\beta_0^{PA} = 2,28347$ (diferença em valor absoluto em torno de 0,00541).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{SS} + \frac{\sigma_2^2}{2} = -0,00164 + \frac{4,7993e-05}{2} = -0,001616$, estimativa mais aproximada para $\beta_2^{PA} = -0,00157$ (diferença em valor absoluto em torno de 0,000046).

Contudo, se considerarmos o modelo GLMM-ar(1), a aproximação do intercepto é dada por $\beta_0^{SS} + \frac{\sigma_0^2}{2} = 2,11070 + \frac{0,34438}{2} = 2,28289$, estimativa bastante próxima de $\beta_0^{PA} = 2,28268$ (diferença em valor absoluto em torno de 0,00021).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{SS} + \frac{\sigma_2^2}{2} = -0,00228 + \frac{3,1396e-10}{2} = -0,00228$, estimativa muito próxima a $\beta_2^{PA} = -0,00225$ (diferença em valor absoluto em torno de 0,00003).

Em ambos os casos, as aproximações entre os parâmetros dos modelos PA e SS mostraram-se bastante próximas. Um resumo dessas aproximações encontra é apresentado na Tabela 7.4.

Tabela 7.4 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão de Poisson com função de ligação *log*

Estrutura de correlação	Parâmetro	β^{PA}	β^{PA*}
exch/cs	Intercepto	2,28347	2,28888
	Period	-0,00157	-0,001616
AR(1)	Intercepto	2,28268	2,28289
	period	-0,00225	-0,00228

Fonte: O autor.

Os comandos do software R para a obtenção dos resultados das Tabelas 7.2 e 7.3 encontram-se na Tabela 7.5.

Tabela 7.5 – Comandos do R para modelos PA e SS de Regressão de Poisson com função de ligação *log*

```
# Chamar a biblioteca simstudy
library(simstudy)

# Criar uma base inicial
def <- defData(varname = "xbase", formula = 10, variance = 0.5,
               dist = "gamma", id = "id")

# Criar fórmula que estabelecerá os parâmetros da Poisson (variável resposta)
def2 <- defDataAdd(varname = "lambda",
                   formula = "xbase + period * Trt",
                   dist="nonrandom",
                   link = "identity")

# Fixar semente (reprodutibilidade)
set.seed(123)

# Simular 10.000 observações (para cada um os 10.000 id's)
dt <- genData(10000, def)

# Designar, para cada um dos 100.00 id's, um tratamento (0 ou 1)
longdef <- trtAssign(dt, nTrt = 2, balanced = TRUE, grpName = "Trt")

# Adicionar 6 períodos (de cada id) na base de dados
longdef1 <- addPeriods(longdef, idvars = "id", nPeriods = 6)

# Adicionar a coluna "lambda" na base de dados
longdef2 <- addColumns(def2, longdef1)

### Gerar a variável resposta NewPois, correlacionada, com 6 medidas repetidas, com
### distribuição de poisson e parâmetros dados pela coluna "lambda".
longdef3 <- addCorGen(dtOld = longdef2, idvar = "id", nvars = 6,
                     rho = .6, corstr = "cs", dist = "poisson",
```

```

param1 = "lambda", cnames = "NewPois")

##### Modelagem GEE
library(geepack)

geefit2 <- geeglm(NewPois ~ factor(Trt) * period,
                  family = poisson(link = "log"),
                  corstr = "exchangeable",
                  data = longdef3,
                  id = id)

# Visualizar a saída do modelo com 5 dígitos
options(digits = 5)
summary(geefit2)

##### Modelagem GLMM
##### Dropping intercept-slope covariance
library(MASS)
library(nlme)

glmm_sim_2 <- glmmPQL(NewPois ~ factor(Trt) * period,
                      random = list(id = pdDiag( ~ period)),
                      family = poisson(link = "log"),
                      data = longdef3)

##### Atualizar o modelo adicionando estrutura de correlação "cs"
glmm_sim_2a <- update(glmm_sim_2, corr = corCompSymm())

options(digits = 5)
summary(glmm_sim_2a)

```

Fonte: O autor.

7.3.2 Relação entre PA e SS via Regressão Logística com Função de Ligação *logit*

Nesta seção, é importante destacar que a base de dados a ser simulada é diferente, apenas no sentido de que a variável resposta passa a ser formada por valores binários (0 e 1). As outras características são, fundamentalmente, as mesmas (correlação e número de períodos).

Com relação aos modelos SS (GLMM-cs e GLMM-ar1) a serem estimados, isto é, de regressão logística com função de ligação *logit*, a expressão geral assume a forma

$$\begin{aligned} \text{logit}(u_{ijt}) = & (\beta_0^{SS} + b_0) + \beta_1^{SS}(\text{Treat}) + (\beta_2^{SS} + b_2)(\text{period}) \\ & + \beta_3^{SS}(\text{Treat} \times \text{period}). \end{aligned} \quad (7.16)$$

As estimativas desses modelos encontram-se na Tabela 7.6. Com base na Tabela 7.6, é possível notar que, ao nível de $\alpha = 0,05$, o intercepto e os coeficientes *Treat* e de interação mostraram-se estatisticamente significantes em ambos os modelos (p-valor < 0,05). Em contrapartida, o coeficiente *period* não se mostrou significativo em nenhum dos modelos (p-valor de 0,223 e 0,1681).

Tabela 7.6 – Estimativas dos modelos SS (regressão logística e função de ligação *logit*)

Método	Parâmetro	Estimativa	SE	t-valor	p-valor
GLMM-cs	Intercepto	-0,39292	0,027360	-14,3610	0,0000
	factor(Treat)1	-0,09897	0,038824	-2,5492	0,0108
	period	-0,00690	0,005665	-1,2186	0,2230
	factor(Treat)1:period	0,03199	0,008019	3,9893	0,0001
	σ_0^2	0,0044102			
	σ_2^2	0,0086269			
GLMM-ar1	Intercepto	-0,45063	0,027036	-16,6676	0,0000
	factor(Treat)1	-0,09105	0,038364	-2,3733	0,0176
	period	-0,01064	0,007716	-1,3785	0,1681
	factor(Treat)1:period	0,03044	0,010930	2,7847	0,0054
	σ_0^2	0,88135			
	σ_2^2	0,018948			

Fonte: O autor.

Os modelos marginais a serem estimados da classe de modelos PA (de regressão logística com estruturas de correlação “*exchangeable*” e “*ar1*” e função de ligação *logit*) são dados por

$$\begin{aligned} \text{logit}(\mu_{ijt}) = & \beta_0^{PA} + \beta_1^{PA}(\text{Treat}) + \beta_2^{PA}(\text{period}) \\ & + \beta_3^{PA}(\text{Treat} \times \text{period}). \end{aligned} \quad (7.17)$$

As estimativas de ambos os modelos PA encontram-se na Tabela 7.7.

Tabela 7.7 – Estimativas dos modelos PA (regressão logística e função de ligação *logit*)

Método	Parâmetro	Estimativa	SE	Wald	Pr(> W)
GEE-exch	Intercepto	-0,41917	0,02483	285,10	< 2e-16
	factor(Treat)1	-0,09218	0,03531	6,82	0,009
	period	-0,00505	0,00507	0,99	0,319
	factor(Treat)1:period	0,02824	0,00715	15,62	7.8e-05
	ϕ	1			
	α	0,462			
GEE-ar1	Intercepto	-0,42258	0,02617	260,69	< 2e-16
	factor(Treat)1	-0,08486	0,03712	5,23	0,0222
	Period	-0,00534	0,00739	0,52	0,4700
	factor(Treat)1:period	0,02731	0,01035	6,96	0,0083
	ϕ	1			
	α	0,536			

Fonte: O autor.

Com base na Tabela 7.7, é possível notar que, ao nível de $\alpha = 0,05$, o intercepto e os coeficientes Treat e de interação mostraram-se estatisticamente significantes em ambos os modelos (GEE-exch e GEE ar1). Apenas o coeficiente *period* não se mostrou estatisticamente significativo em nenhum dos dois modelos (p-valor de 0,319 e 0,47).

As correlações estimadas em ambos os modelos (0,462 e 0,536) ficaram abaixo, ainda que relativamente próximas, do valor estabelecido na simulação (0,6). Em ambos os modelos, os parâmetros de dispersão exibiram as mesmas estimativas, $\phi = 1$, resultados que não sugerem evidências de sobredispersão.

Com relação às aproximações de β_0 e β_2 entre os modelos PA e SS em um contexto de regressão logística e com função de ligação *logit*, a fórmula (28) será utilizada.

Assim, por exemplo, no que diz respeito à aproximação do intercepto entre os modelos PA e SS (modelo GLMM-exch como referência), temos que $\beta_0^{PA*} \approx \frac{-0,39292}{\sqrt{1+0,346 \cdot 0,0044102}} = -0,39262$, tendo ficado relativamente próximo a $\beta_0^{PA} = -0,41917$ (diferença em valor absoluto em torno de 0,02655).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{PA*} \approx \frac{-0,0069}{\sqrt{1+0,346 \cdot 0,0086269}} = -0,0068897$, estimativa bem próxima de $\beta_2^{PA} = -0,00505$ (diferença em valor absoluto em torno de 0,0018397).

Contudo, se considerarmos o modelo GLMM-ar(1), a aproximação do intercepto é dada por $\beta_0^{PA*} \approx \frac{-0,45063}{\sqrt{1+0,346 \cdot 0,88135}} = -0,39448$, estimativa muito próxima a $\beta_0^{PA} = -0,42258$ (diferença em valor absoluto em torno de 0,0281).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{PA*} \approx \frac{-0,01064}{\sqrt{1+0,346 \cdot 0,018948}} = -0,010605$, estimativa relativamente próxima a $\beta_2^{PA} = -0,00534$ (diferença em valor absoluto em torno de 0,005265).

Em ambos os casos, as relações entre os parâmetros dos modelos PA e SS mostraram-se bastante próximas. Um resumo dessas aproximações encontra-se na Tabela 7.8.

Os comandos do R para a obtenção das estimativas das Tabelas 7.7 e 7.6 encontram-se na Tabela 7.9.

Tabela 7.8 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão logística com função de ligação *logit*

Estrutura de correlação	parâmetro	β^{PA}	β^{PA*}
exch/cs	Intercepto	-0,41917	-0,39266
	Period	-0,00505	-0,006891
AR(1)	Intercepto	-0,42258	-0,4002
	period	-0,00534	-0,010609

Fonte: O autor.

Tabela 7.9 – Comandos do R para modelos PA e SS de Regressão Logística com função de ligação *logit*

```
Modelo de regressão logística com função de ligação logit e estruturas "cs" e "exchangeable"

##### Definição da base (xbase)
def <- defData(varname = "xbase", formula = 5, variance = 0.2, dist = "gamma", id = "id")

##### Criação de uma variável p (valores entre 0 e 1) a partir de xbase
def <- defData(def, varname = "p", formula = "-2.0 + 0.3 * xbase",
               dist="nonrandom", link = "logit")

# Fixar semente (garantir reprodutibilidade)
set.seed(123)

# Simular 10.000 observações (para os 10.000 id's)
dt <- genData(10000, def)

# Designar cada um dos 10.000 id's para um dos tratamentos (0 ou 1)
longdef <- trtAssign(dt, nTrt = 2, balanced = TRUE, grpName = "Trt")

# Criar e adicionar variável p2
def2 <- defDataAdd(varname = "p2", formula = "p + 0.1 * p * Trt * 0.1 * period",
                   dist = "nonrandom", link = "identity")

# Adicionar os 6 períodos (de cada id) na base de dados
longdef1 <- addPeriods(longdef, idvars = "id", nPeriods = 6)

# Inserir variável p2 em um contexto longitudinal
dt2 <- addColumns(def2, longdef1)

### Criar variável resposta
longdef3 <- addCorGen(dtold = dt2, idvar = "id", nvars = 6,
                    rho = .6, constr = "cs", dist = "binary",
```

```

param1 = "p2", cnames = "NewBin")

##### Modelagem GEE
library(geepack)

geefit4 <- geeglm(NewBin ~ factor(Trt) * period,
                  data = longdef3,
                  id = id,
                  family = binomial(link = "logit"),
                  corstr = "exchangeable")

options(digits = 5)
summary(geefit4)

##### Modelagem GLMM
library(MASS)
library(nlme)

mod4 <- glmmPQL(NewBin ~ factor(Trt) * period,
                 random = list(id = pdDiag( ~ period)),
                 data = longdef3,
                 family = binomial(link = "logit"))

# Criar uma estrutura de correlação "cs"
# Foi necessário introduzir uma estimativa inicial de correlação 0,3.

cs1 <- corCompSymm(value = 0.3, form = ~ period | id)
cs1 <- Initialize(cs1, data = longdef3)

# Atualizar o modelo com estrutura de correlação "cs"
mod4a <- update(mod4, correlation = corCompSymm(cs1), control = lmeControl(opt = "optim"))

options(digits = 5)
summary(mod4a)

Modelo de regressão logística com função de ligação logit e estrutura "ar1"

# Procedimentos a partir da criação da variável resposta
longdef3 <- addCorGen(dtold = dt2, idvar = "id", nvars = 6,
                     rho = .6, corstr = "ar1", dist = "binary",
                     param1 = "p2", cnames = "NewBin")

##### Modelagem GEE
library(geepack)

geefit3 <- geeglm(NewBin ~ factor(Trt) * period,
                  data = longdef3,
                  id = id,
                  family = binomial(link = "logit"),
                  corstr = "ar1")

options(digits = 5)
summary(geefit3)

```

```
##### Modelagem GLMM

library(MASS)
library(nlme)

mod3 <- glmmPQL(NewBin ~ factor(Trt) * period,
                 random = list(id = pdDiag( ~ period)),
                 data = longdef3,
                 family = binomial(link = "logit"))

##### Modelagem GLMM com estrutura de correlação "ar1"

# Atualizar o modelo com estrutura de correlação "ar1"
mod3a <- update(mod3, correlation = corAR1())

options(digits = 5)
summary(mod3a)
```

Fonte: O autor.

7.3.3 Relação entre PA e SS via Regressão Logística com Função de Ligação *probit*

Devemos ressaltar que a simulação do banco de dados ainda continua sendo feita com função de ligação *logit* (característica do pacote *simstudy*), mas a estimação dos modelos será com a função de ligação *probit*.

A expressão para os modelos SS (GLMM-cs e GLMM-ar1) a serem estimados de regressão logística com função de ligação *probit* é dada por

$$\begin{aligned} \text{probit}(u_{ijt}) = & (\beta_0^{SS} + b_0) + \beta_1^{SS}(\text{Treat}) + (\beta_2^{SS} + b_2)(\text{period}) \\ & + \beta_3^{SS}(\text{Treat} \times \text{period}) \end{aligned} \quad (7.18)$$

As estimativas para os modelos SS dados em (7.18) encontram-se na Tabela 7.10.

Com base na Tabela 7.10, é possível notar que, ao nível de $\alpha = 0,05$, o intercepto e os coeficientes *Treat* e de interação mostraram-se estatisticamente significantes em ambos os modelos (p-valor < 0,05). Por outro lado, o coeficiente *period* não se mostrou significativo em nenhum dos modelos (p-valor de 0,2377 e 0,1435).

Os modelos a serem estimados da classe de modelos PA (de regressão logística com estruturas de correlação “*exchangeable*” e “*ar1*” e função de ligação *probit*) são dados por

$$\begin{aligned} \text{probit}(\mu_{ijt}) = & \beta_0^{PA} + \beta_1^{PA}(\text{Treat}) + \beta_2^{PA}(\text{period}) \\ & + \beta_3^{PA}(\text{Treat} \times \text{period}) \end{aligned} \quad (7.19)$$

As estimativas para os modelos PA dados por (7.19) encontram-se na Tabela 7.11.

Tabela 7.10 – Estimativas dos modelos SS (regressão logística e função de ligação *probit*)

Método	Parâmetro	Estimativa	SE	t-valor	p-valor
GLMM-cs	Intercepto	-0,278674	0,0184259	-15,1241	0,0000
	factor(Treat)1	-0,066646	0,0261209	-2,5514	0,0107
	period	-0,004612	0,0039055	-1,1808	0,2377
	factor(Treat)1:period	0,021200	0,0055269	3,8358	0,0001
	σ_0^2	0,00093948			
	σ_2^2	0,004971			
GLMM-ar1	Intercepto	-0,291787	0,0167311	-17,4397	0,0000
	factor(Treat)1	-0,058595	0,0237158	-2,4707	0,0135
	period	-0,006869	0,0046947	-1,4630	0,1435
	factor(Treat)1:period	0,019535	0,0066471	2,9389	0,0033
	σ_0^2	0,36948			
	σ_2^2	0,0069039			

Fonte: O autor.

Com base na Tabela 7.11, é possível notar que, ao nível de $\alpha = 0,05$, o intercepto e os coeficientes Treat e de interação mostraram-se estatisticamente significantes em ambos os modelos (GEE-exch e GEE ar1). O coeficiente *period* não se mostrou estatisticamente significativo em nenhum dos dois modelos (p-valor de 0,319 e 0,47).

Tabela 7.11 – Estimativas dos modelos PA (regressão logística e função de ligação *probit*)

Método	Parâmetro	Estimativa	SE	Wald	Pr(> W)
GEE-exch	Intercepto	-0,26186	0,01541	288,70	<2e-16
	factor(Treat)1	-0,05714	0,02187	6,82	0,009
	period	-0,00313	0,00315	0,99	0,319
	factor(Treat)1:period	0,01751	0,00443	15,62	7,85e-05
	ϕ	1			
	α	0,462			
GEE-ar1	Intercepto	-0,26397	0,01624	264,07	<2e-16
	factor(Treat)1	-0,05264	0,02300	5,24	0,0221
	period	-0,00331	0,00458	0,52	0,4700
	factor(Treat)1:period	0,01695	0,00642	6,97	0,0083
	ϕ	1			
	α	0,536			

Fonte: O autor.

Com relação às re As correlações estimadas em ambos os modelos exibiram valores idênticos aos modelos logísticos com função de ligação *logit* (0,462 e 0,536), ficando, também, abaixo, ainda que relativamente próximas, do valor estabelecido na simulação (0,6). Em ambos os modelos, os parâmetros de dispersão exibiram, de modo similar aos modelos logísticos com função de ligação *logit*, a mesma estimativa, $\phi = 1$, resultados que não sugerem evidências de sobredispersão.

lações matemáticas das aproximações entre parâmetros dos modelos PA e SS, em um contexto de regressão logística e função de ligação *probit*, a relação dada em (7.10) será utilizada.

Assim, por exemplo, no que diz respeito à aproximação do intercepto entre os modelos PA e SS (modelo GLMM-exch como referência), temos que $\beta_0^{PA*} \approx \frac{-0,278674}{\sqrt{1+0,00093948}} = -0,27854$, tendo ficado relativamente próximo a $\beta_0^{PA} = -0,26186$ (diferença em valor absoluto em torno de 0,01668).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{PA*} \approx \frac{-0,004612}{\sqrt{1+0,004971}} = -0,0046$, estimativa próxima de $\beta_2^{PA} = -0,00313$ (diferença em valor absoluto em torno de 0,00147).

Contudo, se considerarmos o modelo GLMM-ar(1), a aproximação do intercepto é dada por $\beta_0^{PA*} \approx \frac{-0,291787}{\sqrt{1+0,36948}} = -0,24934$, estimativa próxima a $\beta_0^{PA} = -0,26397$ (diferença em valor absoluto em torno de 0,01463).

No que diz respeito ao coeficiente *period*, temos que $\beta_2^{PA*} \approx \frac{-0,006869}{\sqrt{1+0,0069039}} = -0,0068454$, estimativa bem próxima a $\beta_2^{PA} = -0,00331$ (diferença em valor absoluto em torno de 0,0035354).

Em ambos os casos, as relações entre os parâmetros dos modelos PA e SS também se mostraram bastante próximas. Um resumo dessas aproximações encontra-se na Tabela 7.12.

Tabela 7.12 – Análise da aproximação/equivalência entre as estimativas PA e SS – Regressão logística com função de ligação *probit*

Estrutura de correlação	parâmetro	β^{PA}	β^{PA*}
exch/cs	Intercepto	-0,26186	-0,27863
	Period	-0,00313	-0,004608
AR(1)	Intercepto	-0,26397	-0,27475
	period	-0,00331	-0,0068608

Fonte: O autor.

Os comandos para a obtenção das estimativas dispostas nas Tabelas 7.10 e 7.11 encontram-se na Tabela 7.13.

Tabela 7.13 - Comandos do R para modelos PA e SS de Regressão Logística com função de ligação *probit*

Modelo de regressão logística – função de ligação *probit* e estruturas “cs” e “exchangeable”

Definição da base

```
def <- defData(varname = "xbase", formula=5, variance=0.2, dist="gamma",
id="id")
```

OBS: Não foi possível (pelo pacote simstudy) simular variável binária com link *probit*...
!!!

```
def <- defData(def, varname = "p", formula = "-2.0 + 0.3 * xbase",
dist="nonrandom", link = "logit")
```

Fixar semente (garantir reprodutibilidade)

```
set.seed(123)
```

Simular 10.000 observações (para os 10.000 id's)

```
dt <- genData(10000, def)
```

Designar cada um dos 10.000 id's para um dos tratamentos (0 ou 1)

```
longdef <- trtAssign(dt, nTrt = 2, balanced = TRUE, grpName = "Trt")
```

Criar e adicionar variável p2

```
def2 <- defDataAdd(varname = "p2", formula = "p + 0.1 * p * Trt * 0.1 * p
eriod", dist = "nonrandom", link = "identity")
```

Adicionar os 6 períodos (de cada id) na base de dados

```
longdef1 <- addPeriods(longdef, idvars = "id", nPeriods = 6)
```

Inserir variável p2 em um contexto longitudinal

```
dt2 <- addColumns(def2, longdef1)
```

Criar variável resposta

```
longdef3 <- addCorGen(dtOld = dt2, idvar = "id", nvars = 6,
rho = .6, constr = "cs", dist = "binary",
param1 = "p2", cnames = "NewBin")
```

Modelagem GEE

```
library(geepack)
```

```
geefit6 <- geeglm(NewBin ~ factor(Trt) * period,
data = longdef3,
id = id,
family = binomial(link = "probit"),
```



```

corstr = "exchangeable")

options(digits = 5)
summary(geefit6)

##### Modelagem GLMM

library(MASS)
library(nlme)

mod6 <- glmmPQL(NewBin ~ factor(Trt) * period,
               random = list(id = pdDiag( ~ period)),
               data = longdef3,
               family = binomial(link = "probit"))

# Atualizar o modelo com estrutura de correlação "cs"
mod6a <- update(mod6, corr = corCompSymm(), control = lmeControl(opt = "o
ptim"))

options(digits = 5)
summary(mod6a)

Modelo de regressão logística com função de ligação probit e estrutura “ar1”

#### Definição da base

def <- defData(varname = "xbase", formula = 5, variance = 0.2, dist = "ga
mma", id = "id")

#### Criação de uma variável p (valores entre 0 e 1) a partir de xbase

def <- defData(def, varname = "p", formula = "-2.0 + 0.3 * xbase",
               dist="nonrandom", link = "logit")

# Fixar semente (garantir reprodutibilidade)
set.seed(123)

# Simular 10.000 observações (para os 1.000 id's)
dt <- genData(10000, def)

# Designar cada um dos 10.000 id's para um dos tratamentos (0 ou 1)
longdef <- trtAssign(dt, nTrt = 2, balanced = TRUE, grpName = "Trt")

# Criar e adicionar variável p2
def2 <- defDataAdd(varname = "p2", formula = "p + 0.1 * p * Trt * 0.1 * p
eriod", dist = "nonrandom", link = "identity")

# Adicionar os 6 períodos (de cada id) na base de dados
longdef1 <- addPeriods(longdef, idvars = "id", nPeriods = 6)

# Inserir variável p2 em um contexto longitudinal
dt2 <- addColumns(def2, longdef1)

```

```

# Criar variável resposta
longdef3 <- addCorGen(dtOld = dt2, idvar = "id", nvars = 6,
                     rho = .6, corstr = "ar1", dist = "binary",
                     param1 = "p2", cnames = "NewBin")

##### Modelagem GEE
library(geepack)
geefit5 <- geeglm(NewBin ~ factor(Trt) * period,
                  data = longdef3,
                  id = id,
                  family = binomial(link = "probit"),
                  corstr = "ar1")

summary(geefit5)

##### Modelagem GLMM
library(MASS)
library(nlme)
mod5 <- glmmPQL(NewBin ~ factor(Trt) * period,
                 random = list(id = pdDiag( ~ period)),
                 data = longdef3,
                 family = binomial(link = "probit"))

##### Modelagem GLMM com estrutura de correlação "ar1"
mod5a <- update(mod5, corr = corAR1())

options(digits = 5)
summary(mod5a)

```

Fonte: O autor.

7.4 Conclusões

As simulações de dados longitudinais foram realizadas e dois tipos de variável resposta eram de particular interesse: i) dados de contagem (regressão de Poisson) e ii) dados binários (regressão logística), sendo que, para a estimação dos modelos de regressão logística, foram consideradas duas funções de ligação: i) *logit* e ii) *probit*.

Com relação aos resultados gerais dos modelos, o coeficiente associado ao tratamento foi estatisticamente significativo em quase todos os modelos, a não ser pelo modelo de Poisson (GEE-ar1). O coeficiente de interação (entre tratamento e *period*) foi significativo em todos os modelos. No entanto, o coeficiente de *period* não apresentou significância estatística em praticamente nenhum dos modelos estimados.

Em termos da estrutura dos efeitos aleatórios, o modelo em questão era de intercepto e coeficiente aleatórios, e as aproximações entre as estimativas dos modelos PA e SS se mostraram, independentemente das estruturas de correlações utilizadas (“compound symmetry”

ou “ar1”), coerentes com a teoria apresentada, tendo as diferenças (em valores absolutos) permanecidas em patamares baixos.

A despeito de uma base de dados simulada, a condução de um número arbitrariamente elevado de UEs ($N=10.000$) foi realizada com sucesso, ainda que, em alguns casos, um tempo maior para o processamento das iterações tenha sido necessário para a estimação dos modelos SS (GLMM).

8 CONCLUSÕES E LIMITAÇÕES

Com relação aos propósitos da Tese, é possível afirmar que os objetivos foram alcançados.

Uma análise da classe de modelos GEE foi realizada no Capítulo 2. Para isso, os dois artigos que deram origem a essa classe de modelos foram discutidos – Liang e Zeger (1986) propõem uma extensão da classe de modelos GLM para a análise de dados longitudinais, enquanto que Zeger e Liang (1986) sugerem uma abordagem unificadora válida tanto para variáveis contínuas quanto para discretas, via função de quase-verossimilhança. Além disso, foi considerada a proposta de modelo longitudinal de Zeger et al. (1985) (portanto anterior ao surgimento da classe de modelos GEE), restrita para o caso de uma variável resposta binária e que também levava em consideração uma estrutura de correlação para as respostas.

No Capítulo 3, foi introduzida a classe de modelos GLMM, mas, com destaque inicial para a classe de modelos que a antecedeu: LMM. A lógica de seus métodos de estimação também foi apresentada.

No Capítulo 4, a estrutura de um modelo HGLM foi apresentada, tendo, inicialmente, sido introduzida por uma classe de modelos mais restrita, HLM, válida somente para variáveis resposta com distribuição normal.

No Capítulo 5, foi apresentado o software HLM 7, especializado em estimação de modelos hierárquicos. Uma base de dados com características longitudinal e hierárquica foi analisada com o software HLM 7 e também com os softwares SPSS e R. As propriedades do software HLM 7 foram exploradas, e uma análise comparativa entre as estimativas desses softwares foi conduzida.

No Capítulo 6, as classes de modelos PA e SS são examinadas tanto para modelos com respostas binárias quanto para modelos com respostas de Poisson. Algumas relações gerais entre os parâmetros estimados dessas classes de modelos (PA e SS) são avaliadas levando em consideração distintas funções de ligação.

No Capítulo 7, simulações de bancos de dados (com elevados tamanhos amostrais) e estimações de modelos foram conduzidas no intuito de averiguar o grau de aproximação das relações existentes entre as classes de modelos PA e SS, tanto para respostas binárias quanto para respostas com distribuição de Poisson e para as funções de ligação consideradas no Capítulo 6 (*log* para dados de Poisson e *logit* e *probit* para regressão logística). As aproximações entre as estimativas dos modelos PA e SS mostraram-se bastante próximas.

A título de tema(s) para projetos futuros, novas rodadas de simulações e estimações de modelos (PA e SS) poderiam ser empreendidas com modificações (planejadas e controladas) nas características dos dados longitudinais simulados, no intuito de avaliar os efeitos dessas variações no grau de aproximação das estimativas dos parâmetros. Algumas dessas modificações poderiam ser:

- simulação dos dados com número aleatório de observações para cada UE;
- simulação dos dados com um certo grau de assimetria na distribuição das UEs no que diz respeito aos tipos de tratamentos;
- alteração na medida de correlação, ρ , das respostas.

9 REFERÊNCIAS

- Agresti, A. (2015) *Foundations of Linear and Generalized Linear Models*. NJ: Wiley.
- Anderson, D. A., and Aitkin, M. (1985) Variance Components Models with Binary Response: Interviewer Variability. *J. Roy. Statist. Soc.*, (Series B) 47, 203-210.
- Andrade, T. A. N. (2013) *Verossimilhança Perfilada nos Modelos Lineares Generalizados com Sobredispersão*. (Dissertação de Mestrado). Universidade Federal de Pernambuco, Recife.
- Aiken, L. S., Aguinis, H., Appelbaum, M., Boodoo, G. M., Edwards, M. C., Gonzalez, R. D., Panter, A., Park, D., & Patelis, T. (2009) Report of the Task Force for Increasing the Number of Quantitative Psychologists. *American Psychological Association*. Recuperado de: <https://www.apa.org/science/leadership/bsa/quantitative/index.aspx>
- Azzalini, A. (1983) Maximum likelihood estimation of order m for stationary stochastic processes. *Biometrika*, 70(2), 381-387,
- Bates, D. M. (2010) *lme4: Mixed-effects modelling with R*. Recuperado de: <http://lme4.r-forge.r-project.org/book/>
- Breslow, N. E. (2004) *Whither PQL?* Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data. Edited by D. Y. Lin, P. J. Heagerty. Springer.
- Breslow, N. E., & Clayton, D. G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421), 9-25.
- Brown, H., & Prescott, R. (2006) *Applied Mixed Models in Medicine*. 2nd ed. NJ: Wiley.
- Capanu, M., and Begg, C. B. (2011) Hierarchical modeling for estimating relative risks of rare genetic variants: Properties of the pseudo-likelihood method. *Biometrics*, 67(2), 371-380.
- Capanu, M., Gönen, M., & Begg, C. B. (2013) An Assessment of estimation methods for generalized linear mixed models with binary outcomes. *Stat Med.*, 32(26), doi:10.1002/sim.5866.
- Casella, G., & Berger, R. L. (2002) *Statistical Inference*. 2nd ed. CA: Duxbury.
- Chartier, S., & Faulkner, A. (2008) General Linear Models: An Integrated Approach to Statistics. *Tutorial in Quantitative Methods for Psychology*, 4(2), 65-78.
- Corbeil, R. R. and Searle, S. R. (1976) Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. *Technometrics*, 18(1), 31-38.
- Cousineau, D. (2005) The rise of quantitative methods in Psychology. *Tutorial in Quantitative Methods for Psychology*, 1(1), 1-3.

- Cox, R. D. (1958) The Regression analysis of binary sequences (with discussion). *J. Roy. Statist. Soc.*, (Series B) 20, 215-242.
- Cox, R. D. (1970) *The Analysis of Binary Data*. London: Methuen.
- Demidenko, E. (2013) *Mixed models: Theory and applications with R*. 2nd ed. NJ: Wiley.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Roy. Statist. Soc.*, 39(1), 1-38.
- Diggle, P., Liang, K-Y. and Zeger, S. L. (1994) *Longitudinal Data Analysis*. Oxford: Oxford University Press.
- Diggle, P. J., Heagerty, P. J., Liang, K. Y., & Zeger, S. L. (2002) *Analysis of Longitudinal Data*. 2nd ed. UK: Oxford.
- Edgington, E. S. (1964) A tabulation of inferential statistics used in psychology journals. *American Psychologist*, 19, 202-203.
- Ernster et al. (1987) Benign and malignant breast disease: Initial study results of serum and breast fluid analyses of endogenous estrogens. *J. Nat. Canc. Inst.*, 79, 949-960.
- Fitzmaurice, G. & Molenberghs, G. (2009) *Advances in longitudinal data analysis: An historical perspective*. CRC Pres Inc., Boca Raton.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011) *Applied Longitudinal Analysis*. 2nd ed. NJ: Wiley.
- Gelman, A., & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. NY: Cambridge.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2006) *An Introduction to Bayesian Analysis: Theory and Methods*. NY: Springer.
- Gibbons, R. D., Hedeker, D., & Dutoit, S. (2010) Advances in Analysis of Longitudinal Data. *Annu Rev Clin Psychol.*, 27(6), 79-107.
- Graham, J. M. (2008) The General Linear Model as structural equation modeling. *Journal of Educational and Behavioral Statistics*, 3, 485-506.
- Green, P. J. (1987) Penalized Likelihood for General Semi-parametric Regression Models. *International Statistical Review*, 55(3), 245-259.
- Grizzle, J. E., Starmer, C. F. and Koch, G. G. (1969) Analysis of categorical data by linear models. *Biometrics*, 25, 489-504.
- Halekoh, U., Højsgaard, S., & Yan, J. (2006) The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, 15(2).

- Hardin, J. W., & Hilbe, J. M. (2013) *Generalized Estimating Equations*, 2nd Ed. CRC Press Inc., Boca Raton.
- Harville, D. (1976) Extension of the Gauss-Markov Theorem to include the estimation of random effects. *The Annals of Statistics*, 4(2), 384-395.
- Harville, D. A. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, 72(358), 320-338.
- Heagerty, P. J., & Zeger, S. L. (2000) Marginalized Multilevel Models and likelihood Inference. *Statistical Science*, 15(1), 1-26.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014) *Multilevel and Longitudinal Modeling with IBM SPSS*. 2nd ed. USA: Routledge.
- Hedeker, D., Dutoit, S., Demirtas, H.; & Gibbons, R. D. (2018) A Note on Marginalization of Regression Parameters from Mixed Models of Binary Outcomes. *Biometrics*, 74(1), 354-361.
- Hedeker, D., & Gibbons, R. D. (2006) *Longitudinal Data Analysis*. NJ: John Wiley & Sons, Inc.
- Henderson, C. R. (1950) Estimation of genetic parameters (abstract). *Ann. Math. Statist.*, 21, 309-310.
- Hogg, R. V., McKean, J. W., & Craig, A. T. (2005) *Introduction to Mathematical Statistics*. 6th ed. USA: Pearson.
- JANG, J. J. (2011) *Working correlation selection in generalized estimating Equations*. (Tese de doutorado). University of Iowa. Recuperado de: <https://ir.uiowa.edu/etd/2719/>
- Jiang, J. (2007) *Linear and Generalized Linear Mixed Models and Their Applications*. USA: Springer.
- Johnson, N. L. and Kotz, S. (1970) *Distributions in Statistics: Continuous Univariate Distributions*, 2, Bostons: Houghton-Mifflin.
- Kauermann, G., & Carroll, R. J. (2001) A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96, 1387-1396.
- Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehmann, R. B. (1971) A general methodology for the analysis of repeated measurements of categorical data. *Biometrics*, 33, 133-158.

- Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H. and Lehnen R. G. (1977) A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33, 133-158.
- Knox, T. A., Spiegelman, D., Skinner, S. C. et al. (2000) Diarrhea and Abnormalities of gastrointestinal function in a cohort of men and women with HIV Infection. *American Journal of Gastroenterology*, 95, 3482-89.
- Laird, N. M. (1982) Empirical Bayes estimates using the nonparametric maximum likelihood estimate for the prior. *Journal of Statistical Computation and Simulation*, 15, 211-220.
- Laird, N. M. and Ware, J. H. (1982) Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963-974.
- Liang, K. Y., & Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.
- Lindley, D. V.; & Smith, A. F. M. (1972) Bayes Estimates for the Linear Model. *J. Roy. Statist. Soc., Series B (Methodological)*, 34(1), 1-41.
- Long, J. D., Loeber, R., & Farrington, D. P. (2009) Marginal and Random Intercepts for Longitudinal Binary Data With Examples From Criminology. *Multivariate Behav. Res.*, 44(1), 28-58.
- Mccullagh, P., and Nelder, J. A. (1983) *Generalized Linear Models*. 2nd Ed. London. New York: Chapman and Hall.
- Mccullagh, P., & Nelder, J. A. (1989) *Generalized Linear Models*. 2nd ed. USA: Chapman and Hall.
- Mcculloch, C. E., & Searle, S. R. (2001) *Generalized, Linear and Mixed Models*. NJ: Willey.
- Mclachlan, G. J., & Thriyambakam, K. (2008) *The EM Algorithm and Extensions*. 2nd ed. NJ: Wiley.
- Molenberghs, G., & Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. NJ: Springer.
- Murray, D. M., Hannan, P. J., Wolfinger, R. D., Baker, W. L., and Dwyer, J. H. (1998) Analysis of data from group-randomized trials with repeat observations on the same groups. *Statistics in Medicine*, 17, 1581-600.
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface*, 14, 20170213.
- Nelder, J., & Wedderburn, R. W. M. (1972) Generalized Linear Models. *J. Roy. Statist. Soc.*, 135(3), 370-384.

- Neuhaus, J. M., and Jewell, N. (1993) A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrics*, 80, 807-815.
- Neuhaus, J. M., Kalbfleisch, J. D., & Hauck, W. W. (1991) A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data. *International Statistical Review*, 59(1), 25-35.
- Ogata, Y. (1980) Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC. *J. Appl. Prob.*, 17, 59-72.
- Oichi, Y. and Prentice, R. L. (1984) Likelihood inference in correlated probit regression. *Biometrika*, 71, 531-543.
- Paula, G. A. (2013) *Modelos de Regressão com apoio computacional*. Instituto de Matemática e Estatística. Universidade de São Paulo. Recuperado de: https://www.ime.usp.br/~giapaula/texto_2013.pdf
- Patterson, H. D. and Thompson, R. (1971) Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3), 545-554.
- Pawitan, Y. (2001) *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. NY: Oxford University Press.
- Pinheiro, J. C., & Bates, D. M. (2000) *Mixed Effects Models in S and S-Plus*. NY: Springer-Verlag.
- R CORE TEAM (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de: <https://www.R-project.org/>.
- Raudenbush, S. W (2009) Targets of inference in hierarchical models for longitudinal data. *Longitudinal Data Analysis*. Edited by Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. USA: Chapman & Hall/CRC.
- Raudenbush, S. W., & Bryk, A. S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. USA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Condon, R. T. C. Jr., & Toit, M. du. (2011) *HLM 7: Hierarchical Linear and Nonlinear Modeling*. USA: Scientific Software International. Recuperado de: <http://www.ssicentral.com>.
- Ritz, J., & Spiegelman, D. (2004) Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research*, 13, 309-323.

- Self, S. G. and Liang, K-Y. (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, 82(398), 605-610.
- Skidmore, S. T., & Thompson, B. (2010) Statistical Techniques Used in Published Articles: A Historical Review of Reviews. *Educational and Psychological Measurement*, 70(5), 777-795.
- Smith, A. F, M, (1973) Bayes Estimates for the Linear Model. *J. Roy. Statist. Soc., Series B* (Methodological), 34(1), 1-41.
- Stiratelli, R., Laird, N. and Ware, J. H. (1984) Random-effects models for serial observations with binary response. *Biometrics*, 40, 961-971.
- Stram, D. O. and Lee, J. W. (1994) Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, 50(4), 1171-1177.
- Sun, L. (2011) *Comparison of Different Estimation Methods for Linear Mixed Models and Generalized Linear Mixed Models*. Supervisor: Lars Rönnegard. D-level Essay in Statistics. Dalarna University, Sweden. Recuperado de: <https://pdfs.semanticscholar.org/9416/64ee6283e248c0f433fe6495acb033e961ac.pdf>
- Szmaragd, C., Clarke, P., & Steele, F. (2013) Subject specific and population average models for binary longitudinal data: a tutorial. *Longitudinal and life course studies*, 42(2), 147-165.
- Thompson, B. (2006) *Foundations of behavioral statistics: An insight-based approach*. NY: Guilford Press.
- Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. NY: Springer,
- Vonesh, E. F., & Chinchilli, V. M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. NY: Marcel Dekker.
- Wedderburn, R. W. M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, 61(3), 439-447.
- Wolfinger, R., & O'connell, M. (1993) Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *J. Statist. Comput. Simul.*, 48, 233-243.
- Woltman, H., Feldstain, A., Mackay, J. C., & Rocchi, M. (2012) An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69.
- Young, M. L., Preisser, J. S., Qaqish, B. F., & Wolfson, M. (2007) Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials. *Statistical Methods in Medical Research*, 16, 167-184.

- Zeger, S. L., & Liang, K. Y. (1986) Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42(1), 121-130.
- Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988) Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44(4), 1049-1060.
- Zeger, S. L., Liang, K. Y., & Self, S. G. (1985) The analysis of binary longitudinal data with independent covariates. *Biometrika*, 72(1), 31-38.
- Zhang, H., Yu, Q., Feng, C., Gunzler, D., Wu, P., & Tu, X. M. (2012) A new look at the difference between the GEE and GLMM when modeling longitudinal count responses. *Journal of Applied Statistics*, 39(9), 2067-2079.
- Ziegler, A. (2011) *Generalized Estimating Equations*. NY: Springer.