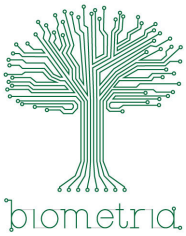
Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Métodos de estimação do excesso de mortes durante a pandemia de COVID-19 no Brasil

Fernando Andrade

Botucatu
2025

Fernando Andrade

Métodos de estimação do excesso de mortes durante a pandemia de COVID-19 no Brasil

Dissertação de Mestrado apresentada ao Curso de Programa de Pós-graduação em Biometria da Universidade Estadual Paulista "Júlio de Mesquita Filho" como parte dos requisitos necessários para a obtenção do título de Mestre em Biometria.

Orientador: Profa. Dra. Luzia Aparecida Trinca

Botucatu
2025

Andrade, F. A, Fernando.

Métodos de estimação do excesso de mortes durante a pandemia de COVID-19 no Brasil/ Fernando Andrade. –, 2025-
51 p. 1 :il. (colors; grafs; tabs).

Orientador: Profa. Dra. Luzia Aparecida Trinca

Dissertação de Mestrado – Universidade Estadual Paulista "Júlio de Mesquita Filho".
Instituto de Biociências – Câmpus de Botucatu.
Programa de Pós-graduação em Biometria.

1. Modelo Generalizado Misto. 2. Predição. 2. COVID-19. I. Profa. Dra. Luzia Aparecida Trinca. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". III. Título

Fernando Andrade

Métodos de estimação do excesso de mortes durante a pandemia de COVID-19 no Brasil

Dissertação de Mestrado defendido e aprovado em Botucatu, Fevereiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Luzia Aparecida Trinca
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Orientador

Profa. Dra. Mariana Cúri
Universidade de São Paulo
Examinador

Profa. Dra. Liciana Vaz de Arruda Silveira
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Examinador

Dedico este trabalho aos meus pais, Ana Paula e Fabio, por todo o amor, carinho, compreensão e ensinamentos que sempre me ofereceram, não apenas durante minha jornada acadêmica, mas ao longo da vida. À minha orientadora, Luzia Aparecida Trinca, por sua preocupação genuína e sábia orientação ao longo da produção deste trabalho. Sua dedicação à pesquisa é exemplar, e sem ela, não teria sido possível completar esta dissertação. E, também, aos meus amigos, tanto aqueles presentes em minha rotina em Botucatu quanto os de minha cidade de origem. Em especial, dedico este trabalho a Gustavo, Glória, Gabriel, Lucas, Carol, Abrantes, Thiago e Victória, pelas conversas, momentos especiais e por sempre me incentivarem a conquistar meus objetivos.

Agradecimentos

A realização deste trabalho foi possível graças ao apoio e incentivo de diversas pessoas e instituições, às quais expresso minha mais profunda gratidão.

Agradeço, primeiramente, à minha orientadora, Luzia Aparecida Trinca, pela orientação precisa, paciência e dedicação ao longo de todo o desenvolvimento desta dissertação. Suas valiosas contribuições e disponibilidade foram fundamentais para que este trabalho alcançasse seus objetivos.

À CAPES, pela concessão da bolsa de estudos, que possibilitou a dedicação necessária às atividades de pesquisa. Reconheço a importância do apoio financeiro para a continuidade e avanço das ciências no Brasil.

Aos meus pais, Ana Paula e Fabio, pela educação e valores que me transmitiram, pelo amor incondicional e pelo apoio constante em todas as fases da minha vida. Este trabalho é também uma conquista de vocês.

Aos meus amigos, por estarem ao meu lado em momentos de dificuldade e celebração. Suas palavras de incentivo, conversas e risadas me ajudaram a superar os desafios e seguir em frente com determinação.

Agradeço também a todos do Programa de Pós-Graduação em Biometria da Universidade Estadual Paulista "Júlio de Mesquita Filho", em Botucatu, pelo ambiente acadêmico acolhedor e pelas valiosas trocas de conhecimento ao longo desta jornada.

Por fim, a todos aqueles que, de alguma forma, contribuíram para que este momento fosse possível, deixo aqui o meu mais sincero agradecimento.

Essentially, all models are wrong, but some are useful.
George Box

Resumo

A análise da mortalidade de eventos significativos, tais como pandemias, é um estudo fundamental para compreender a saúde local e avaliar seu impacto na população. Um estudo do impacto causado é de grande relevância para avaliar os possíveis fatores que poderiam ter sido mitigados e também para o planejamento de respostas a futuros eventos adversos. Neste contexto, uma maneira de executar esse estudo é através da estimação do excesso de mortes ocorrido no período. Para isso, deve-se levar em conta a diferença entre a mortalidade observada e a mortalidade que seria esperada na não ocorrência da pandemia, que é predita por modelos estatísticos. O grande desafio é a escolha de métodos de predição que levam em consideração a incerteza envolvida no processo. Dentro da bibliografia de predições, os modelos de efeitos mistos destacam-se como métodos de complexidade mediana. Este trabalho buscou explorar as vantagens e dificuldades associadas aos Modelos Lineares Generalizados Mistos (MLGM) para predizer as mortes na situação de não ocorrência da pandemia de COVID-19. Os modelos Gaussiano e Binomial Negativa apresentaram resultados semelhantes na estimativa da mortalidade esperada, ambos mostrando-se adequados para o contexto estudado. Além disso, este trabalho utilizou simulações como alternativa aos métodos numéricos mais sofisticados presentes na literatura de MLGM, o que permitiu calcular predições e incertezas associadas de forma eficiente. Tais resultados contribuem para o avanço na modelagem de dados correlacionados e heterogêneos, além de fornecer ferramentas úteis para o planejamento de ações em saúde pública.

Palavras-chave: Modelos Lineares, Modelos Lineares Generalizados, Modelos Mistos, Modelos Generalizados Mistos, COVID-19, Modelagem Estatística, Estimação, Predição, Simulação.

Abstract

The analysis of mortality from significant events, such as pandemics, is fundamental to understanding local health and assessing its impact on the population. Examining the impact caused is highly relevant for evaluating factors that could have been mitigated and for planning responses to future adverse events. In this context, one way to conduct this study is by estimating the excess deaths that occurred during the period. This requires considering the difference between observed mortality and the mortality that would have been expected in the absence of the pandemic, which is predicted by statistical models. The main challenge lies in choosing prediction methods which offer the precision or uncertainty associated.. Within the prediction literature, mixed-effects models stand out as mid-complexity methods. This study aimed to explore the advantages and challenges associated with Generalized Linear Mixed Models (GLMM) in predicting deaths under the scenario of the COVID-19 pandemic's non-occurrence. The Gaussian and Negative Binomial models produced similar results in estimating the expected mortality, both proving adequate for the context studied. Moreover, this study employed simulations as an alternative to the more sophisticated numerical methods found in the GLMM literature, enabling efficient calculation of predictions and associated uncertainties. These findings contribute to advancing the modeling of correlated and heterogeneous data and provide useful tools for planning public health actions.

Keywords: Linear Models, Generalized Linear Models, Mixed Models, Generalized Mixed Models, COVID-19, Statistical Modeling, Estimation, Prediction, Simulation.

Lista de figuras

Figura 1 – Comportamento periódico da mortalidade brasileira a partir da 1^a semana de 2015 até a 9^a semana de 2020, juntamente com os modelos mistos ajustados. 33

Figura 2 – Valores marginais ajustados *versus* Resíduos marginais padronizados (esquerda) e Valores condicionais ajustados *versus* Resíduos condicionais padronizados (direita) para o Modelo 2. 35

Figura 3 – Índice de Lesaffre e Verbeke (esquerda) e QQ-χ^2_q da distância de Mahalanobis, $\mathcal{M}_i$ (direita) para o Modelo 2. 35

Figura 4 – QQ-Normal dos resíduos minimamente confundidos padronizados para o Modelo 2. 36

Figura 5 – Gráfico QQ-Normal dos resíduos tipo Deviance (esquerda) e Gráfico seminormal dos resíduos tipo Deviance (direita) para o Modelo Poisson Misto. . 38

Figura 6 – Resíduos de Pearson *versus* Valores ajustados (esquerda) e Resíduos de trabalho (*working*) *versus* Valores ajustados (direita) para o Modelo Poisson Misto. 38

Figura 7 – QQ-Plot dos resíduos tipo Deviance (esquerda) e Gráfico Seminormal dos resíduos tipo Deviance (direita) para o Modelo Binomial Negativa Misto. . . 40

Figura 8 – Resíduos *versus* Valores ajustados (esquerda) e Resíduos (*Working*) *versus* Valores ajustados (direita) para o Modelo Binomial Negativa Misto. 40

Figura 9 – Estimativas e intervalos de predição a 95% (Modelo Normal Misto), para a mortalidade esperada (faixa verde), mortalidade esperada + mortalidade por COVID-19 no período pandêmico (faixa vermelha) e mortalidade observada (linha vermelha) para o ano de 2020 (superior) e de 2021 (inferior). 42

Figura 10 – Estimativas e intervalos de predição a 95% (Modelo Binomial Negativa Misto), para a mortalidade esperada (faixa azul), mortalidade esperada + mortalidade por COVID-19 no período pandêmico (faixa vermelha) e mortalidade observada (linha vermelha) para o ano de 2020 (superior) e de 2021 (inferior). 43

Figura 11 – Excesso de mortes utilizando os modelos mistos Normal (faixa pontilhada verde) e Binomial Negativa (faixa pontilhada em azul) e as mortes observadas por COVID-19 durante o ano de 2020. 45

Lista de tabelas

Tabela 1 – Uso das estatísticas específicas para diagnóstico de ajuste de um Modelo Linear Misto. 26
Tabela 2 – Estatísticas globais da qualidade de ajuste dos modelos Gaussinos 34
Tabela 3 – Estimativas dos parâmetros do Modelo Normal Misto e respectivas significâncias para os efeitos fixos 36
Tabela 4 – Estatísticas globais da qualidade de ajuste dos modelos Poisson 37
Tabela 5 – Estimativas dos parâmetros do Modelo Poisson Misto, e respectivas significâncias para os efeitos fixos 37
Tabela 6 – Estatísticas globais da qualidade de ajuste dos modelos Binomial Negativa. . 39
Tabela 7 – Estimativas dos parâmetros do Modelo Binomial Negativa Misto, e respectivas significâncias para os efeitos fixos 39

Lista de abreviaturas e siglas

AIC	Critério de Informação de Akaike
ANOVA	Análise de Variância
BIC	Critério de Informação Bayesiano
BLUE	Melhor estimador linear não-viesado
BLUP	Melhor preditor linear não-viesado
DO	Documento de Óbito
EBLUE	Melhor estimador linear não-viesado empírico
EBLUP	Melhor preditor linear não-viesado empírico
EM	Maximização de Expectativa
GEE	Equações de estimação generalizadas
GLS	Mínimos Quadrados Generalizados
IBGE	Instituto Brasileiro de Geografia e Estatística
IRWLS	Mínimos Quadrados Reponderados
LR	Razão de Verossimilhanças
ML	Máxima Verossimilhança
MLG	Modelo Linear Generalizado
MLGM	Modelo Linear Generalizado Misto
MLM	Modelo Linear Misto
OLS	Mínimos Quadrados Ordinários
PQL	Quase-Verosimilhança Penalizada
REML	Máxima Verossimilhança Restrita
SIM	Sistema de Informação sobre Mortalidade
SVS	Secretaria de Vigilância em Saúde
US CDC	Centros de Controle e Prevenção de Doenças dos Estados Unidos
WHO	Organização Mundial de Saúde

Sumário

1 INTRODUÇÃO 1
1.1 Objetivos 2
1.2 Esboço da tese 2

2 FUNDAMENTAÇÃO TEÓRICA 4
2.1 Modelos Lineares e Lineares Generalizados 4
2.2 Modelos de Efeitos Mistos 7
2.2.1 Modelo Linear Misto 8
2.2.1.1 Modelos Marginal e Condicional 10
2.2.1.2 Estimação dos Parâmetros 11
2.2.2 Modelos Lineares Generalizados Mistos 14
2.2.2.1 Sobredispersão 15
2.2.2.2 Modelos Marginal e Condicional 16
2.2.2.3 Estimação dos Parâmetros 17
2.2.2.3.1 Aproximação de Laplace 19
2.2.2.3.2 Quadratura Gaussiana 20
2.2.3 Inferência 20
2.2.4 Diagnósticos do Ajuste 24
2.2.4.1 Técnicas para MLM 24
2.2.4.2 Técnicas para MLGM 26
2.2.5 Predições para Novas Observações 27

3 MATERIAL E METODOLOGIA 30
3.1 Metodologia 30
3.2 Material 32

4 RESULTADOS 33
4.1 Ajustes dos Modelos 33
4.1.1 Modelo Gaussiano 34
4.1.2 Modelo de Poisson 36
4.1.3 Modelo Binomial Negativa 39
4.2 Estimação da Mortalidade Esperada e Excesso de Mortes no Período Pandêmico 41

5 CONCLUSÃO 46

Referências **48**

1 Introdução

A análise da mortalidade é essencial para compreender o impacto de pandemias e avaliar a saúde populacional. A epidemia de COVID-19 no Brasil levantou questões cruciais sobre o número de mortes e sua magnitude, influenciada pela gestão precária da saúde pública nos anos de 2020 e 2021 (MALTA et al., 2021), resultando, até agosto de 2023, em mais de 700 mil mortes causadas diretamente pela doença (Ministério da Saúde, 2023b).

Entretanto, uma pandemia da magnitude da COVID-19 ou outros eventos adversos podem influenciar a mortalidade por diversos fatores indiretos. Exemplos incluem dificuldades de acesso a cuidados e tratamentos de outras doenças, aumento de estresse e ansiedade que agravam a saúde mental, e dificuldades financeiras que impactam as condições de moradia e alimentação. Por outro lado, a redução de mortes por outras causas, como doenças infecciosas e acidentes, devido à menor circulação pública, também deve ser considerada.

Uma avaliação abrangente do impacto da pandemia sobre a mortalidade, considerando fatores diretos e indiretos, é essencial para entender a experiência no país e identificar aspectos que poderiam ter sido mitigados. Além disso, esses dados são valiosos para o planejamento de respostas a futuros eventos adversos. Uma maneira de estudar esse impacto é através da estimação do excesso de mortalidade no período.

O excesso de mortalidade devido a um evento adverso é definido como a diferença entre a mortalidade observada e aquela que seria esperada sob condições normais, ou seja, na ausência do evento (WHO, 2023). A mortalidade observada é registrada pelos órgãos oficiais, e o grande desafio está em prever qual seria a mortalidade se a pandemia não tivesse ocorrido. A literatura sobre métodos de estimação é ampla. Ela inclui desde abordagens simples, como médias históricas, até modelos mais sofisticados que requerem dados censitários da população. Entre os métodos de complexidade mediana estão os modelos lineares de efeitos mistos (COLONIA et al., 2023).

Este trabalho explora as vantagens e desafios da aplicação de Modelos Lineares Generalizados Mistos (MLGM) para a predição da mortalidade esperada no Brasil, caso a pandemia não tivesse ocorrido. Com base nas predições, o foco foi estimar o excesso de mortes durante a pandemia de COVID-19, visto que os MLGMs são ferramentas adequadas para lidar com dados correlacionados e heterogêneos, como o número de mortes ao longo do tempo.

Os modelos escolhidos foram: Normal, Poisson e Binomial Negativa. O modelo Gaussiano (Normal) tem vasta literatura abordando aspectos de modelagem, como estimação, diagnóstico e predição. Já os modelos Poisson e Binomial Negativa, comumente utilizados para dados de contagem, têm uma literatura menos extensa, o que gera desafios na predição de mortalidade para os anos pandêmicos.

Entretanto, esses desafios foram superados com técnicas de simulação. No modelo Gaussiano, as predições para valores futuros podem ser calculadas teoricamente, assim como sua incerteza. A aplicação de métodos simulacionais para os demais modelos também trouxe resultados satisfatórios.

Entre os modelos ajustados, o modelo de Poisson, devido à sua insensibilidade com a sobredispersão, não apresentou bom ajuste. No entanto, a correção pela inclusão de uma fonte extra de variabilidade no modelo Binomial Negativa trouxe bons resultados. Assim, os modelos mais parcimoniosos foram o Normal e o Binomial Negativa, que foram utilizados para calcular o excesso de mortalidade.

1.1 Objetivos

Este trabalho tem como objetivo geral aplicar versões do MLGM para estimar o excesso de mortalidade no Brasil durante a pandemia de COVID-19. As versões incluíram: o modelo que assume normalidade dos dados (Gaussiano), Poisson e Binomial Negativa, sendo estes últimos comumente utilizados para dados de contagem. Além disso, o trabalho buscou extrair informações sobre as vantagens e dificuldades de cada estratégia de modelagem. Entre os objetivos específicos, o trabalho dispõe-se:

1. Estimar a mortalidade esperada no período pandêmico com base nos modelos escolhidos;
2. Comparar o ajuste dos modelos Normal, Poisson e Binomial Negativa, avaliando suas vantagens e limitações;
3. Analisar o excesso de mortalidade através das predições ajustadas e seus intervalos de incerteza;
4. Identificar os fatores associados às estimativas de excesso de morte.

1.2 Esboço da tese

Além desta introdução (Capítulo 1), o trabalho está organizado da seguinte forma: o Capítulo 2 (Fundamentação Teórica) descreve os conceitos básicos de modelagem estatística, com ênfase nos modelos lineares e generalizados, detalhando os modelos mistos, suas especificações, métodos de estimação, técnicas de diagnóstico e predições. A metodologia é apresentada no

Capítulo 3, detalhando o processo de ajuste dos modelos, as ferramentas computacionais utilizadas e a fonte dos dados.

O Capítulo 4 apresenta os resultados do estudo, como os ajustes dos modelos, seus diagnósticos, as predições de mortalidade para 2020 sob a hipótese de não ocorrência da pandemia, com suas respectivas incertezas calculadas via simulação, e, por fim, o excesso de mortes. Para finalizar, o Capítulo 5 contém a conclusão do projeto, da qual há uma breve retomada dos objetivos iniciais e os principais resultados discutidos durante os demais capítulos.

Essa estrutura busca apresentar de maneira lógica e detalhada as bases teóricas, metodológicas e empíricas necessárias para avaliar o impacto da pandemia de COVID-19 na mortalidade no Brasil, contribuindo para estudos futuros em modelagem estatística e planejamento de políticas públicas.

2 Fundamentação Teórica

2.1 Modelos Lineares e Lineares Generalizados

Um modelo estatístico é uma coleção de premissas com estruturas que permitem efetuar estimações de quantidades populacionais de interesse ao utilizar dados coletados e, por exemplo, obter predições de valores futuros (FOX; WEISBERG, 2011). As quantidades de interesse são parâmetros que podem representar efeitos de variáveis ou fatores explanatórios sobre uma ou mais variável aleatória de interesse e/ou representar como essas variáveis estão relacionadas. Usualmente a variável aleatória é representada por Y e as explanatórias por X_j ($j = 1,\ 2,\ \ldots,\ p$). Dentro das classes dos modelos estatísticos, os modelos de regressão formam uma ampla classe de modelos, na qual os lineares são os de menor complexidade e de grande utilidade quando Y é do tipo contínua e razoavelmente bem comportada.

Os modelos lineares clássicos envolvem uma parte fixa preditiva, $\mathbb{E}(Y)$, responsável pela variabilidade sistemática, e uma parte aleatória, responsável pela variabilidade não sistemática de Y. Para n realizações de Y e p parâmetros na parte preditiva, na forma matricial, o modelo linear pode ser escrito como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

em que $\mathbf{Y}$ e $\boldsymbol{\epsilon}$ são vetores colunas $(n \times 1)$, $\boldsymbol{\beta}$ é o vetor de parâmetros fixos $(p \times 1)$, $\mathbf{X}$ é a matriz de especificação do modelo $(n \times p)$, juntamente com as seguintes premissas:

1. $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$.
2. $\mathbb{C}(\boldsymbol{\epsilon}) = \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{I}$.
3. A matriz $\mathbf{X}$ é de posto completo, isto é, $rk(\mathbf{X}) = p < n$.
4. $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$.

Veja, que sob tais condições, tem-se $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ e $\mathbb{V}(\mathbf{Y}) = \mathbb{C}(\mathbf{Y}) = \sigma^2\mathbf{I}$, tal que $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. O número total de parâmetros, incluindo σ^2, é $p + 1$, com uma característica notável, a independência entre os parâmetros de média e o de variância, que, nesse caso, é pressuposto constante. Implicitamente, esse modelo se encaixa no plano de coleta de dados por amostragem aleatória simples ou por experimento inteiramente aleatorizado, cuja variável

resposta é contínua e, pelo menos, aproximadamente normal, seja na escala original ou após alguma transformação.

Entre as técnicas de estimação dos parâmetros na abordagem clássica, a mais utilizada é o Método dos Mínimos Quadrados dos erros (OLS), que consiste em encontrar $\boldsymbol{\beta}$ que minimiza a soma $\text{SQ}_E = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$, resultando, no caso de $\mathbf{X}'\mathbf{X}$ invertível, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. As propriedades deste estimador são bem conhecidas, incluindo não tendenciosidade e normalidade. O Teorema de Gauss-Markov diz que o estimados OLS é uma ótima escolha, visto que dentre todas as opções não enviesadas de funções lineares de $\mathbf{Y}$, o método OLS oferece o estimador de variância mínima (FARAWAY, 2014) dada por $\mathbb{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Além do mais, sob normalidade dos erros, o estimador via o método de Máxima Verossimilhança (ML) coincide com $\hat{\boldsymbol{\beta}}$.

No entanto, em muitas situações, a variável resposta pode ser contínua assimétrica, discreta ou mesmo categórica, de forma que a aproximação normal não é apropriada. Ademais, outras premissas importantes do modelo e que limitam sua aplicabilidade, são a homoscedasticidade e a inexistência de correlação entre os erros.

Os Modelos Lineares Generalizados (MLG) foram desenvolvidos por Nelder e Wedderburn (1972) de forma a estender os modelos lineares de modo que a variável resposta não precise seguir a distribuição gaussiana, mas bastando pertencer à Família Exponencial. Desta maneira, seja $\mathbf{Y}$ um vetor de variáveis aleatórias independentes tal que cada Y_i possui densidade dada por

$$f_Y(y_i|\theta_i, \phi) = \exp\left[\phi(y_i\theta_i - b(\theta_i)) + c(y_i, \phi)\right],$$

em que, para todo i, $\mathbb{E}(Y_i) = \mu_i = b'(\theta)$ e $\mathbb{V}(Y_i) = \phi^{-1}\mathcal{V}(\mu_i)$, donde $\mathcal{V}_i = \mathcal{V}(\mu_i) = \mathrm{d}\mu_i/\mathrm{d}\theta_i$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão ou precisão. A função de variância exerce um papel fundamental na família exponencial, posto que caracteriza como a média e a variabilidade da variável aleatória se relacionam. Em particular, para Y uma variável discreta com valores não-negativos, refletindo contagens de ocorrência de algum evento, se $\mathcal{V}_i = \mathcal{V}(\mu_i) = \mu$, com $\mu > 0$, então a distribuição associada é a Poisson. Para verificarmos isso, veja que se $Y \sim Po(\mu),\ \mu > 0$, temos que

$$f_Y(y|\mu) = \frac{e^{-\mu}\mu^y}{y!} = \exp\left(y\ln\left(\mu\right) - \mu - \ln\left(y!\right)\right).$$

Logo, $\theta = \ln(\mu), b(\theta) = \mu = \exp(\theta)$. Portanto, $\mathcal{V}(\mu_i) = \mathrm{d}\mu_i/\mathrm{d}\theta = \frac{\mathrm{d}\exp(\theta)}{\mathrm{d}\theta} = \exp(\theta) = \mu$. Ademais, há a parte sistemática

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

em que $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, $p < n$, é o vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ representa o vetor de valores de variáveis explicativas, de acordo com a especificação do modelo preditivo e $g(\cdot)$ é uma função monótona e diferenciável, denominada por função de ligação (PAULA, 2013). Assim como no modelo linear clássico, o termo "preditor linear" indica que a linearidade é com relação à forma que os parâmetros $\boldsymbol{\beta}$ aparecem na sua expressão e a função de ligação não é necessariamente linear.

Em resumo, os modelos lineares generalizados são definidos por três partes:

1. A distribuição de Y_i, que deve pertencer à família exponencial.
2. A parte sistemática: $\eta_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$.
3. A função de ligação: $g(\mu_i) = \eta_i$.

Observa-se que o modelo de regressão linear clássico, que pressupõe a normalidade da variável resposta, é um caso particular do MLG. Algumas outras distribuições populares abrangidas são: Gama, Log-Normal, Poisson, Bernoulli e Binomial, para as quais, os parâmetros de média e variância são dependentes.

A estimação dos parâmetros, sob o paradigma de inferência frequencista, é baseada no princípio da máxima verossimilhança. Para uma amostra simples de Y de tamanho n, denotada por $y_1, \ldots, y_n$, o logaritmo da função de verossimilhança é dado por

$$\ln L(\boldsymbol{\theta}, \phi|\mathbf{y}) = \sum_{i=1}^{n} [\phi(y_i\theta_i - b(\theta_i)) + c(y_i, \phi)]. \tag{2.1}$$

Assim, derivando-se (2.1) com relação à $\boldsymbol{\beta}$ e igualando a zero, obtém-se as soluções, isto é, o estimador do parâmetro $\boldsymbol{\beta}$. A dificuldade da obtenção das soluções é a determinação de uma solução explicita, que ocorre apenas para a situação em que os dados seguem a distribuição Gaussiana, que utiliza a função de ligação identidade. Para os demais casos, a maximização de (2.1) é por meio do processo de Newton-Raphson juntamente do Scoring de Fisher. Entretanto, o método dos mínimos quadrados iterativos re-ponderados (IRWLS) apresenta solução idêntica. A ideia é linearizar a função de ligação aplicada em y, $g(y)$ por meio da expansão de Taylor:

$$g(y) \approx g(\mu) + (y - \mu) \left.\frac{\mathrm{d}g(y)}{\mathrm{d}y}\right|_{y=\mu} = g(\mu) + (y - \mu)\frac{\mathrm{d}g(\mu)}{\mathrm{d}\mu} = z$$

em que z é denominada variável de trabalho. Na notação matricial, tem-se

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{-1/2}\mathbf{V}^{-1/2}(\mathbf{Y} - \mu)$$

tal que $\mathbb{E}(\mathbf{Z}) = g(\mu(\mathbf{X})) = \mathbf{X}\boldsymbol{\beta}$ e variância $\mathbb{V}(\mathbf{Z}) \approx \phi \times \mathbf{W}$. Desta maneira, $\mathbf{Z}$ funciona como uma variável resposta modificada e $\mathbf{W}$ é a matriz de pesos que se alteram a cada passo do processo iterativo dado por:

1. Entre com valores iniciais para $\hat{\mu}_0$ e obtenha $g(\hat{\mu}_0)$;
2. Calcule $\hat{z}_0$;
3. Calcule os pesos $\hat{w}_0^{-1}$;
4. Estime $\boldsymbol{\beta}$ por mínimos quadrados ponderados e obtenha $\hat{g}$;

5. Itere os passos 2-4 até a convergência.

Logo, a soma a ser minimizada no passo 4, via notação matricial, é

$$(\hat{\mathbf{Z}} - \mathbf{X}\boldsymbol{\beta})\widehat{\mathbf{W}}(\hat{\mathbf{Z}} - \mathbf{X}\boldsymbol{\beta})$$

com $\widehat{\mathbf{W}} = \text{diag}(w_1, \ldots, w_n)$, tal que o operador diag cria uma matriz diagonal dos pesos $w_i = \left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i}\right)^2 \frac{1}{\mathcal{V}_i}$. Consequentemente,

$$\widehat{\mathbb{V}(\hat{\boldsymbol{\beta}})} \approx (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \times \hat{\phi}.$$

Para mais detalhes, aconselha-se a leitura de autores consagrados no tema como Paula (2013), Fahrmeir et al. (2013), Fitzmaurice, Laird e Ware (2012), Faraway (2016).

Os modelos descritos previamente, conhecidos como de efeitos fixos, não incorporam possíveis agrupamentos no processo de coleta dos dados passíveis de contribuir com heterogeneidade na variabilidade e/ou correlação nas observações da variáveis resposta. Contudo, eles podem ser estendidos através da incorporação de fatores associados a efeitos aleatórios, permitindo a modelagem tanto do ponto de vista médio ou populacional, quanto do ponto de vista específico de grupos ou indivíduos, cada um com múltiplas observações. O modelo linear que possui ambos os tipos de efeito é denominado por Modelo Linear Misto (MLM). Extensões similares aos MLGs, resultando no modelos lineares generalizados mistos (MLGM), e aos modelos não lineares, juntamente com os lineares mistos formam a classe dos Modelos Mistos (PINHEIRO; BATES, 2009).

2.2 Modelos de Efeitos Mistos

Modelos Mistos são amplamente utilizados para descrever relações entre a variável resposta e algumas covariáveis cujas observações são agrupadas seguindo pelo menos um fator de classificação. Alguns exemplos de dados agrupados são: dados longitudinais, coletados no espaço ou tempo, dados com medidas repetidas, dados com estrutura hierárquica ou multinível e dados experimentais nos quais há restrições na aleatorização como, por exemplo, nos delineamentos em blocos e delineamentos em múltiplos estratos (tipo parcela subdividida e extensões). Tais dados, além de incorporarem variabilidade extra, devido aos termos aleatórios, compartilham uma característica em comum, uma estrutura de correlação entre as observações de um mesmo agrupamento.

Nesse contexto, a modelagem inclui duas classes de parâmetros, os efeitos fixos que modelam da média de $\mathbf{Y}$ e os efeitos aleatórios que modelam a estrutura de covariância-variância de $\mathbf{Y}$, tornando os modelos mistos uma ferramenta flexível e adequada para análise de dados correlacionados (MCCULLOCH; SEARLE; NEUHAUS, 2011). Em particular, os parâmetros relacionados à variabilidade extra são chamados de componentes de variância.

Assim como o Modelo Linear Gaussiano é um caso particular de Modelo Linear Generalizado, segue que sob certas suposições a respeito dos efeitos aleatórios, o Modelo Linear Misto é uma caso particular de Modelo Linear Generalizado Misto, embora, do ponto de vista de modelagem da estrutura de variabilidade e de correlação, o caso Gaussiano permite maior flexibilidade. Além disso, há diferentes níveis de dificuldades quanto aos métodos de estimação, diagnóstico de ajuste e inferências entre essas duas classes de modelos, justificando a apresentação que segue, em seções separadas.

2.2.1 Modelo Linear Misto

Fisher, em seu livro de 1925, já começava a esboçar uma maneira para estimar componentes de variância, utilizando o método de Análise de Variâncias (ANOVA), que iguala cada quadrado médio ao seu valor esperado, ou seja, utilizando o método dos momentos em dados balanceados. Yates (1940) e Henderson (1953) estenderam o método de Fisher para dados não balanceados, contudo, implicando em estimativas que podem não ser únicas. Hartley e Rao (1967) mostraram que soluções únicas podem ser obtidas através do método de máxima verossimilhança, embora, conhecidamente, resultando em estimadores de componentes de variância enviesados, mesmo nos delineamentos mais simples. Cinco anos depois, Patterson e Thompson (1971) desenvolveram o método da máxima verossimilhança restrita ou residual (REML), atualmente preferido por ser não enviesado quando os dados são balanceados. Para dados não balanceados, o método é iterativo, o que impossibilita demonstrar sua não viciosidade, mas espera-se que tenha melhores propriedades em relação ao estimador de máxima verossimilhança.

Mas apenas com Laird e Ware (1982), os pioneiros a formularem os Modelos Lineares Mistos para dados longitudinais em dois estágios, a metodologia de estimação dos componentes de variância-covariância se solidificou, mesmo no caso de dados faltantes, através da aplicação do algoritmo EM (*Expectation-Maximization*).

Considerando $\mathbf{Y}_i$ o vetor resposta n_i-dimensional, a formulação dos Modelos Mistos para um estrato de agrupamento é dada por

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n \\ \mathbf{b}_i &\sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_i), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{R}_i) \end{aligned} \tag{2.2}$$

em que n é o número de indivíduos ou grupos, cada um de tamanho n_i, $\boldsymbol{\beta}$ é o vetor p-dimensional de efeitos fixos, $\mathbf{b}_i$ é o vetor q-dimensional de efeitos aleatórios, $\mathbf{X}_i$ $(n_i \times p)$ e $\mathbf{Z}_i$ $(n_i \times q)$ são as matrizes conhecidas do modelo, respectivamente, de efeitos fixos e de efeitos aleatórios e $\boldsymbol{\epsilon}_i$ é o vetor n_i-dimensional dos erros intragrupo com distribuição Normal Multivariada. Complexidades na estrutura de variâncias e covariâncias são ditadas pelos padrões definidos por $\boldsymbol{\Sigma}_i$ e $\mathbf{R}_i$. Por exemplo, $\mathbf{R}_i$ poderá ser modelada a partir de funções de variância não-constantes e possivelmente, covariâncias não nulas. Entretanto, o modelo básico pressupõe $\mathbf{R}_i = \sigma^2\mathbf{I}$ e $\boldsymbol{\Sigma}_i = \sigma_b^2\mathbf{I}$. Ademais, outras premissas são $\mathbf{b}_i \perp \mathbf{b}_{i'}$ e $\mathbf{b}_i \perp \boldsymbol{\epsilon}_i$, isto é, assume-se independência entre efeitos aleatórios

de grupos distintos, assim como independência entre termos aleatórios de níveis distintos intra-grupo.

Os efeitos aleatórios $\mathbf{b}_i$ são definidos para possuírem média $\mathbf{0}$ e, portanto, quaisquer termo de média não nulo deverá ser expresso juntamente com os termos de efeitos fixos. Portanto, as colunas da matriz $\mathbf{Z}_i$ são, geralmente, subconjuntos das colunas de $\mathbf{X}_i$. Assim, o vetor de efeitos aleatórios é completamente caracterizado por sua matriz de variância-covariância $\mathbf{\Sigma}$, que é simétrica e positiva semi-definida (autovalores não negativos).

Com a intenção de facilitar o cálculo computacional, Bates et al. (2015) aponta a importância de expressar a matriz de variância-covariância $\mathbf{\Sigma}$ através da decomposição de Cholesky, dada por

$$\sigma^2\mathbf{\Sigma} = \sigma^2\mathbf{\Delta}\mathbf{\Delta}' \tag{2.3}$$

em que $\mathbf{\Delta}$ é denominado por fator de covariância relativa de $\mathbf{b}_i$.

A formulação apresentada em (2.2) pode ser estendida para múltiplos níveis de agrupamento. Pinheiro e Bates (2009) formula os modelos mistos para dois níveis de estrato de agrupamento aninhados da seguinte maneira: considere $\mathbf{Y}_{ij}$ o vetor de respostas n_{ij}-dimensional, com $i = 1, \ldots, n$ e $j = 1, \ldots, n_i$, em que n é o número de grupos no primeiro nível de estrato e n_i, o número de grupos no segundo nível de estrato dentro do grupo i de primeiro nível, cada um de tamanho n_{ij}; $\mathbf{X}_{ij}$ $(n_{ij} \times p)$ a matriz de especificação dos efeitos fixos, $\mathbf{b}_i$ q_1-dimensional e $\mathbf{b}_{ij}$ q_2-dimensional, respectivamente, os vetores de efeitos aleatórios no primeiro nível e no segundo nível de estratificação, ambos especificados, nesta ordem, com as matrizes modelos dos efeitos aleatórios $\mathbf{Z}_{i,j}$ $(n_{ij} \times q_1)$ e $\mathbf{Z}_{ij}$ $(n_{ij} \times q_2)$. Desta maneira, tem-se

$$\begin{gathered}\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1), \quad \mathbf{b}_{ij} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_2), \quad \boldsymbol{\epsilon}_{ij} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})\end{gathered} \tag{2.4}$$

$\mathbf{b}_i$'s são assumidos serem independentes para diferentes i's, $\mathbf{b}_{ij}$'s são assumidos serem independentes para diferentes i's e j's e independentes de $\mathbf{b}_i$ e, por fim, $\boldsymbol{\epsilon}_{ij}$'s são independentes para diferentes i's ou j's e, também, dos efeitos aleatórios $\boldsymbol{\epsilon}_{ij}$.

Extensões da formulação para um número arbitrário S de níveis de efeitos aleatórios seguem o mesmo padrão anterior, sendo possível generalizar a equação do vetor $\mathbf{Y}_i$ como segue:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{s=1}^{S} \mathbf{Z}_{is}\mathbf{b}_{is} + \boldsymbol{\epsilon}_i \tag{2.5}$$

com $\mathbf{b}_i = (\mathbf{b}_{i1}, \ldots, \mathbf{b}_{iS})' \sim \mathcal{N}_Q(\mathbf{0}, \mathbf{\Sigma}_i)$, de dimensão $Q = \sum_{s=1}^{S} q_s$ e independente de $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{R})$.

As estruturas concernentes aos estratos também podem ser do tipos cruzadas, bastando cuidado na especificação das matrizes $\mathbf{R}$ para gerar o modelo correto, seja com aninhamento e ou cruzamento. Assim, o modelo geral, para o vetor $\mathbf{Y}$ completo, pode ser escrito na forma matricial dada por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}. \tag{2.6}$$

2.2.1.1 Modelos Marginal e Condicional

O modelo definido na expressão (2.6) envolvem as variáveis aleatórias $\mathbf{b}$ e $\mathbf{Y}$ de naturezas distintas. Enquanto $\mathbf{Y}$ é a variável resposta observável, $\mathbf{b}$ é composta por variáveis latentes. No intuito de manter a notação simplificada, na sequência, vamos considerar a configuração de dados longitudinais. As variáveis aleatórias são descritas por suas funções densidades de probabilidades, essenciais para a formulação dos modelos mistos. No primeiro estágio, tem-se a distribuição condicional da variável aleatória dependente $\mathbf{Y}$ dado que os efeitos aleatórios são supostos conhecidos. No segundo estágio tem-se a distribuição (incondicional) dos efeitos aleatórios $\mathbf{b}$.

A distribuição condicional de $\mathbf{Y}_i$ dado $\mathbf{b}_i$, $f_{Y|b}(\mathbf{y}_i|\mathbf{b}_i)$, é uma distribuição normal multivariada com vetor de médias e matriz de variância-covariância dadas, respectivamente, por:

$$\mathbb{E}(\mathbf{Y}_i|\mathbf{b}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i \quad \text{e}$$
$$\mathbb{V}(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{R}_i.$$

Quando $\mathbf{R}_i = \sigma^2\mathbf{I}$, o modelo é dito ser de independência condicional. Detalhando para a observação j do indivíduo i, tem-se $\mathbb{E}(Y_{ij}|\mathbf{b}_i) = \mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$ em que $\mathbf{x}_{ij} = (x^1_{ij}, \cdots, x^p_{ij})'$, $z_{ij} = (z^1_{ij}, \cdots, z^q_{ij})'$ são vetores colunas, que contém os valores das variáveis explanatórias da parte fixa e da parte aleatória para a j-ésima observação do i-ésimo grupo. Portanto, a esperança condicional para uma observação é uma combinação linear dos vetores $\mathbf{x}_{ij}$ e $\mathbf{z}_{ij}$ e realiza predições específicas para o indivíduo i na condição j.

As distribuições (incondicionais) dos efeitos aleatórios, $f_b(\mathbf{b})$, já foram definidas na expressão (2.2) e, pela definição de distribuição condicional, a distribuição conjunta de $\mathbf{Y}_i$ e $\mathbf{b}_i$, $f_{Y,b}(\mathbf{y}_i, \mathbf{b}_i) = f_{Y|b}(\mathbf{y}_i|\mathbf{b}_i)f_b(\mathbf{b}_i)$, ou seja, o produto da distribuição incondicional dos efeitos aleatórios $\mathbf{b}$ e da distribuição condicional de $\mathbf{Y}$ dado $\mathbf{b}$, ambas distribuições gaussianas multivariadas. Portanto, a distribuição conjunta também será gaussiana multivariada.

A distribuição marginal de $\mathbf{Y}_i$, $f_Y(\mathbf{y}_i)$, também é de interesse pois permite a realização de predições a nível populacional, ou seja, em termos médios. Ela é obtida integrando-se a distribuição conjunta de $\mathbf{Y}_i$ e $\mathbf{b}_i$ com relação aos efeitos aleatórios, isto é,

$$f_Y(\mathbf{y}_i) = \int f_Y(\mathbf{y}_i, \mathbf{b}_i)\mathrm{d}\mathbf{b}_i = \int f_{Y|b}(\mathbf{y}_i|\mathbf{b}_i)f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i \tag{2.7}$$

Como $f_{Y,b}$ e f_b são densidades de normais multivariadas, segue que $f_Y(\mathbf{y}_i)$ também é uma normal multivariada. Assim, obtemos

$$\mathbb{E}(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta} \quad \text{e} \tag{2.8}$$
$$\mathbb{V}(\mathbf{Y}_i) = \mathbf{V} = \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}'_i + \mathbf{R}_i. \tag{2.9}$$

As expressões (2.8) e (2.9) implicam que $\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta},\ \mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}'_i + \mathbf{R}_i)$. Observa-se que a matriz de variância-covariância é composta por dois componentes de variabilidade e correlação: $\mathbf{Z}_i\boldsymbol{\Sigma}\mathbf{Z}'_i$ devido aos efeitos aleatórios $\mathbf{b}_i$'s e $\mathbf{R}_i$ em virtude aos termos dos erros $\boldsymbol{\epsilon}_i$'s.

2.2.1.2 Estimação dos Parâmetros

Sejam $\boldsymbol{\beta}$, os parâmetros de média e $\boldsymbol{\theta}$ os parâmetros que modelam a estrutura de variância e covariância. No caso mais simples, $\boldsymbol{\theta} = (\sigma_b^2, \sigma^2)$, a variabilidade entre indivíduos e a variabilidade intra-indivíduos. A função de verossimilhança é uma função de $(\boldsymbol{\beta},\ \boldsymbol{\theta})$ dados os valores observados das variáveis que são observáveis, ou seja, $\mathbf{y}$. Como $\mathbf{b}_i$ são variáveis latentes, o método da Máxima Verossimilhança maximiza a verossimilhança marginal.

Para Y_{ij}, $j = 1, \cdots, n_i$ e $i = 1, \cdots, n$, a função de verossimilhança (L) é o produto das partes:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} L_i(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}_i).$$

Os estimadores de Máxima Verossimilhança de $\boldsymbol{\beta}$ e de $\boldsymbol{\theta}$ são, respectivamente, os vetores $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\theta}}$ que maximizam $\ell = \log L$, isto é, maximizam o logaritmo da função de verossimilhança:

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \log L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{M} \log L_i(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}_i) \\
&= \sum_{i=1}^{n} \log \left(\frac{1}{(2\pi)^{n_i/2}|\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \right) \right) \\
&= -\frac{\sum_{i=1}^{n} n_i}{2} \log(2\pi) - \frac{1}{2}\sum_{i=1}^{n} \log|\mathbf{V}| - \frac{1}{2}\sum_{i=1}^{n} (\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}).
\end{aligned} \tag{2.10}$$

Soluções analíticas existem apenas para modelos e estrutura de dados bem simples e, no geral, a maximização é via algum método iterativo. Contudo, pela expressão acima, sabe-se que, dado $\boldsymbol{\theta}$ conhecido, o estimador de ML de $\boldsymbol{\beta}$ é equivalente ao estimador de mínimos quadrados generalizados (GLS), logo:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}^{-1}\mathbf{Y}_i \right). \tag{2.11}$$

Nesse caso, pode-se demonstrar que,

- $\hat{\boldsymbol{\beta}}_{GLS}$ é não-viesado, independente da escolha de $\boldsymbol{\theta}$, implicando que $\hat{\boldsymbol{\beta}}_{OLS}$ também é não-viesado para $\boldsymbol{\beta}$;
- $\mathbb{V}(\hat{\boldsymbol{\beta}}_{GLS}) = (\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{V}^{-1}\mathbf{X}_i)^{-1}$;
- para amostras grandes, $\hat{\boldsymbol{\beta}}_{GLS}$ tem distribuição Normal, mesmo que $\mathbf{Y}$ não seja Normal.

Grandes amostras significam $n \to \infty$ para n_i's fixos, ou seja, replicações de observações dentro dos grupos são irrelevantes. Embora $\mathbb{E}(\hat{\boldsymbol{\beta}}_{GLS}) = \boldsymbol{\beta}$ para qualquer que seja $\boldsymbol{\theta}$, o uso da estrutura de variância-covariância correta fornece o estimador mais preciso.

Na prática, $\boldsymbol{\theta}$ é desconhecido, e os parâmetros podem ser estimados pelo seguinte algoritmo:

1. Obter $\tilde{\boldsymbol{\beta}}$ usando (2.11) para $\boldsymbol{\theta}$ qualquer;

2. Substituir $\tilde{\boldsymbol{\beta}}$ em (2.10), resultando na chamada função de verossimilhança perfilada,

$$\ell(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{n} n_i}{2} \log(2\pi) - \frac{1}{2}\sum_{i=1}^{n} \log|\mathbf{V}| - \frac{1}{2}\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}) \quad (2.12)$$

   e maximize em $\boldsymbol{\theta}$, tendo em vista que $\mathbf{V}$ é uma função de $\boldsymbol{\theta}$, utilizando algum método iterativo tais como: Newton-Raphson, Escore de Fisher, Gauss-Newton. Com a convergência, obtém-se o estimador $\hat{\boldsymbol{\theta}}_{ML}$.

3. Substituir $\hat{\boldsymbol{\theta}}_{ML}$ em (2.11) e, consequentemente, obtenha o estimador $\hat{\boldsymbol{\beta}}$.

Uma vez obtidos $\hat{\boldsymbol{\beta}}$, denominado por EBLUE (*Empirical Best Linear Unbiased Estimator*) e $\hat{\boldsymbol{\theta}}$, é possível predizer os efeitos aleatórios $\mathbf{b}_i$ através da distribuição condicional

$$\mathbf{b}_i|\mathbf{y} \sim \mathcal{N}(\sigma_b^2\mathbf{Z}_i'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), (\mathbf{Z}_i'\mathbf{Z}_i + \sigma_b^{-2})^{-1}).$$

A média de $\mathbf{b}_i|\mathbf{y}$ é o BLUP (*Best Linear Unbiased Predictor*) de $\mathbf{b}_i$, que substituindo σ_b^2 e $\boldsymbol{\beta}$ por suas estimativas resulta no EBLUP (*Empirical Best Linear Unbieased Predictor*)

$$\tilde{\mathbf{b}}_i = \hat{\sigma}_b^2\mathbf{Z}_i'\hat{\mathbf{V}}^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}).$$

A resposta média para o i-ésimo indivíduo ou grupo é

$$\tilde{\mathbf{y}}_i|\tilde{\mathbf{b}}_i = \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\tilde{\mathbf{b}}_i.$$

Como já brevemente mencionado, mesmo para situações simples, sabe-se que o estimador $\hat{\boldsymbol{\theta}}_{ML}$ é viciado. Outro método de estimação popular é a Máxima Verossimilhança Restrita (REML) proposto por Patterson e Thompson (1971) com a ideia de separar a informação nos dados de forma que uma parte se destina à estimação de $\boldsymbol{\beta}$ e a outra, para a estimação dos parâmetros de variância-covariância em $\boldsymbol{\theta}$.

Desta forma, estima-se os parâmetros de variância apenas com a informação relevante para tal, ou seja, elimina-se a parte referente à $\boldsymbol{\beta}$ da verossimilhança de forma que ela se torne uma função apenas dos parâmetros de variância.

O método utilizado para prosseguir com essa ideia é aplicar uma transformação linear e ortogonal aos dados, $\mathcal{R} = \mathbf{LY}$, tal que a distribuição de $\mathcal{R}$ não dependa de $\boldsymbol{\beta}$. Com isso, basta escolher $\mathbf{L}$ tal que $\mathbb{E}(\mathcal{R}) = \mathbf{0}$. Uma opção, por exemplo, é

$$\mathcal{R} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{I} - \mathbf{H})\mathbf{Y},$$

no qual $\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ e $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é a conhecida matriz chapéu (*hat*) ou matriz de projeção. Assim, o posto de $\mathbf{L} = (\mathbf{I} - \mathbf{H})$ é $\sum_{i=1}^{n} n_i - p$ e a informação nos dados é separada em p contrastes para estimar $\boldsymbol{\beta}$ e $\sum_{i=1}^{n} n_i - p$ para estimar $\boldsymbol{\theta}$. O estimador REML para $\boldsymbol{\theta}$ maximiza

$$\ell_{REML}(\boldsymbol{\theta}) = -\frac{1}{2}\left(\sum_{i=1}^{n} \log|\mathbf{V}| + \sum_{i=1}^{n} \mathbf{r}_i'\mathbf{V}^{-1}\mathbf{r}_i + \log\left|\sum_{i=1}^{n} \mathbf{X}_i\mathbf{V}^{-1}\mathbf{X}_i\right|\right), \quad (2.13)$$

tal que $\mathbf{r}_i$ é a parte observada, referente ao grupo i, do vetor $\mathcal{R}$. Similarmente ao método ML, a maximização de (2.13) é, em geral, por meio de algum método iterativo, que na convergência, fornece $\hat{\boldsymbol{\theta}}_R$, as estimativas REML de $\boldsymbol{\theta}$. Sendo assim, uma vez substituindo $\hat{\boldsymbol{\theta}}_R$ em (2.11), obtém-se $\hat{\boldsymbol{\beta}}_R$, as estimativas REML dos efeitos fixos.

Vale a pena observar que a função em (2.13) não envolve $\boldsymbol{\beta}$, com isso é estranho referir-se a $\hat{\boldsymbol{\beta}}_R$ como estimador REML, mesmo assim, a nomenclatura é popular, indicando que as estimativas foram obtidas utilizando-se $\hat{\boldsymbol{\theta}}_R$. Portanto, comparações entre modelos que se diferenciam em termos de efeitos fixos $\boldsymbol{\beta}$ não podem ser feitas através de valores de verossimilhanças restritas ou critérios que se apropriam dessa função. A recomendação é reajustar o modelo através do método ML para comparação de modelos quanto aos efeitos fixos.

Para dados balanceados e modelos com estrutura de variância-covariância razoavelmente simples, é possível mostrar que os estimadores REML são não-viesados. Para dados desbalanceados e/ou com estruturas mais complexas de variância-covariância, o processo é exclusivamente iterativo e não é possível demonstrar a não tendenciosidade, embora, em teoria, assintoticamente, não há vícios (SEARLE; CASELLA; MCCULLOCH, 2006).

As distribuições dos estimadores tanto por ML, quanto por REML, podem ser obtidas através de aproximações:

1. Assintótica: para número grande de grupos independentes, os estimadores LM e REML convergem para a normalidade;
2. Via método de reamostragem (*bootstrap*);
3. Método Delta: aproxima os momentos da variável aleatória por expansão em série de Taylor.

Os resultados assintóticos são:

$$\hat{\boldsymbol{\beta}} \overset{n\to\infty}{\sim} \mathcal{N}\left(\boldsymbol{\beta}, \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i\right)^{-1}\right)$$

e

$$\hat{\boldsymbol{\theta}} \overset{n\to\infty}{\sim} \mathcal{N}(\boldsymbol{\theta}, \mathcal{I}(\boldsymbol{\theta})^{-1})$$

em que $\mathcal{I}(\boldsymbol{\theta})$ é a matriz de informação empírica, calculada com base na verossimilhança perfilada, dada por

$$\mathcal{I}(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}.$$

Os valores estimados de $\boldsymbol{\theta}$ são substituídos em $\mathbf{V}$ e $\mathcal{I}(\boldsymbol{\theta})$ para obtenção de estimativas destas matrizes. Nota-se que essas estimativas são aproximações de primeira ordem.

2.2.2 Modelos Lineares Generalizados Mistos

Estruturas de agrupamento e correlação entre observações também podem estar presentes quando a variável resposta é discreta ou não Normal, levando à necessidade de inclusão de efeitos aleatórios na equação do preditor linear e gerando a classe dos Modelos Lineares Mistos Generalizados (MLGM), extremamente útil na modelagem de dados binários e contagens longitudinais.

Os MLGM, na literatura, também podem serem vistos como uma extensão dos MLG com a inclusão de efeitos aleatórios, assim como os MLM, uma extensão do Modelo Linear Clássico. É comum encontrar nomes como Modelos Hierárquicos, Modelos Multiníveis e Modelos de Coeficientes Aleatórios como sinônimos dos MLGMs e/ou MLMs (TUERLINCKX et al., 2006).

A formulação de um MLGM é semelhante ao MLG. Utilizando a definição de Fitzmaurice, Laird e Ware (2012) segue que:

1. A distribuição condicional de Y_{ij}, dado o vetor $\mathbf{b}_i$ de efeitos aleatórios, deve pertencer à família exponencial, e, similarmente ao MLG, temos $\mathbb{V}(Y_{ij}|\mathbf{b}_i) = \phi\mathcal{V}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)$, com $\mathcal{V}$ sendo a função de variância, que caracteriza a distribuição de probabilidade da variável aleatória e ϕ^{-1} o parâmetro de dispersão ou escala.

2. Assume-se que a média condicional de $\mathbf{Y}_i|\mathbf{b}_i$ depende tanto dos efeitos fixos, quanto dos efeitos aleatórios por meio de um preditor linear dado por

$$g(\mathbb{E}(\mathbf{Y}_i|\mathbf{b_i})) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

   tal que

$$\mathbb{E}(\mathbf{Y}_i|\mathbf{b_i}) = g^{-1}(\boldsymbol{\eta}_i) = g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)$$

   e $g(\cdot)$ é a função de ligação, monótona e diferenciável.

3. Os efeitos aleatórios possuem alguma distribuição de probabilidade. Em princípio, qualquer distribuição multivariada pode ser assumida para $\mathbf{b}_i$, entretanto, na prática, é comum a utilização da Normal Multivariada com média zero e matriz de covariância de dimensão $q \times q$, $\boldsymbol{\Sigma}$ e $\mathbf{b}_i$'s são independentes das covariáveis X_i's.

Como no caso dos modelos de efeitos fixos, a verificação de que, sob a normalidade dos efeitos aleatórios e de $Y|\mathbf{b}$ e função de ligação identidade, o modelo misto se reduz ao caso já apresentado em detalhes (MLM).

Ilustrando para Y_{ij} uma variável aleatória de contagem relativa ao grupo i, ao longo do tempo, um MLGM poderia ser especificado como

1. Y_{ij}'s, dados os efeitos aleatórios $\mathbf{b}_i$, são independentes e seguem a distribuição de Poisson com $\mathbb{V}(Y_{ij}|\mathbf{b}_i) = \mathbb{E}(Y_{ij}|\mathbf{b}_i)$, em que $\phi = 1$;

2. A média condicional de Y_{ij}, que depende dos efeitos aleatórios, é dada por meio do seguinte preditor linear
$$g(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i.$$
A ligação $g(\cdot) = \log(\cdot)$ implica na seguinte média condicional
$$\mathbb{E}(Y_{ij}|\mathbf{b}_i) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i).$$
Em particular, se o preditor inclui apenas intercepto e efeito de uma variável explanatória t e $\mathbf{x}_{ij} = \mathbf{z}_{ij} = (1,\ t_{ij})'$, tem-se
$$\log(\mathbb{E}(Y_{ij})|\mathbf{b}_i) = \eta_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij},$$
e
$$\mathbb{E}(Y_{ij}|\mathbf{b}_i) = \exp(\beta_0 + b_{0i})\exp((\beta_1 + b_{1i})t_{ij}),$$
o modelo conhecido como de intercepto e inclinação aleatórios, .

3. O vetor de efeitos aleatórios, para esse modelo particular, $\mathbf{b}_i = (b_{0i},\ b_{1i})'$, possue uma distribuição normal bivariada com média zero e matriz de variância-covariância $\boldsymbol{\Sigma}$ de dimensão 2×2.

Como, para a variável Poisson, o parâmetro média é a taxa de ocorrência do evento, seja por intervalo de tempo, de espaço ou qualquer outra escala, de forma que as contagens podem ser vistas como respostas à exposição a um determinado fator, e o preditor linear modela essa taxa, no caso de haver exposições de intensidades distintas, é recomendado incluir essa informação na modelagem. Por exemplo, seja $\frac{\mathbb{E}(Y_{ij}|\mathbf{b}_i)}{u_{ij}}$ a taxa do número de casos de uma doença em uma determinada população em relação à medida de exposição u_{ij}. Logo, para $g(\cdot) = \log(\cdot)$, tem-se
$$\log\left(\frac{\mathbb{E}(Y_{ij}|\mathbf{b}_i)}{u_{ij}}\right) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$$
tal que $\log(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i + \log(u_{ij})$ implicando na seguinte média condicional
$$\mathbb{E}(Y_{ij}|\mathbf{b}_i) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)u_{ij}.$$
O termo $\log(u_{ij})$ é denominado por variável *offset*.

2.2.2.1 Sobredispersão

A utilização de modelos de Poisson para dados de contagem é um procedimento comumente utilizado. Entretanto, em muitas aplicações, os dados apresentam uma variabilidade maior do que aquela que o modelo tem a capacidade de lidar, ou seja, que pode ser predita pelo modelo de Poisson. Em particular, negligenciar a sobredispersão acarreta em subestimação dos erros padrões resultando em inferências inapropriadas.

Entre as técnicas mais utilizadas para lidar com a sobredispersão presente no modelo de Poisson, segundo Fitzmaurice, Laird e Ware (2012), é considerar uma fonte extra de variabilidade no modelo de Poisson

$$\log \mathbb{E}(Y_{ij}|\mathbf{b}_i, \mathrm{e}_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i + \log u_{ij} + \mathrm{e}_{ij}$$

tal que $\log(u_i)$ é a variável *offset* e e_{ij} é a fonte extra de variabilidade em que $\mathrm{e}_{ij} \sim \Gamma(1, \theta)$. É possível mostrar que o modelo em questão é tal que Y_{ij} segue a distribuição Binomial Negativa com média condicional igual a

$$\mathbb{E}(Y_{ij}|\mathbf{b}_i) = \exp\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i\right)u_{ij}$$

ou seja, a mesma média condicional do modelo de Poisson, no entanto, com variância condicional dada por

$$\mathbb{V}(Y_{ij}|\mathbf{b}_i) = \mathbb{E}(Y_{ij}|\mathbf{b}_i) + \phi(\mathbb{E}(Y_{ij}|\mathbf{b}_i))^2 \tag{2.14}$$

em que ϕ é denominado por parâmetro de dispersão. Assim, a modelagem para a variável contagem incorpora uma variabilidade extra não contemplada pela distribuição Poisson.

2.2.2.2 Modelos Marginal e Condicional

A construção de um MLGM já especifica a distribuição condicional de $Y_{ij}|\mathbf{b}_i$ e, portanto, a obtenção de seus parâmetros é automática pela aplicação de $g^{-1}(\cdot)$ e da relação entre média e variância. Por outro lado, a distribuição marginal de Y_{ij}, facilmente obtida no MLM, exige cálculos mais complexos, usualmente sem solução analítica.

Similarmente ao desenvolvimento apresentado na seção 2.2.1.1, conceitualmente, o modelo marginal pode ser obtido integrando, com respeito ao vetor de efeitos aleatórios $\mathbf{b}_i$, o produto das funções densidade $f_Y(Y_{ij}|\mathbf{b}_i)$ e $f_b(\mathbf{b}_i)$, ou seja,

$$f_Y(Y_{ij}) = \int f_Y(Y_{ij}|\mathbf{b}_i)f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i$$

Para fim didático, vamos retomar o caso particular em que é ajustado um modelo Gaussiano com a função de ligação sendo a identidade, e demonstrar a expressão explícita para o modelo marginal, inicialndo com a esperança condicional:

$$g(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i = \mathbb{E}(Y_{ij}|\mathbf{b}_i) = \mathbb{E}(g(Y_{ij}|\mathbf{b}_i)) \Rightarrow g^{-1}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$$

e, consequentemente, aplicando resultados conhecidos à esperança condicional, temos

$$\begin{aligned}\mathbb{E}(Y_{ij}) = \mathbb{E}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) &= \int g^{-1}(\mathbb{E}(Y_{ij}|\mathbf{b}_i))f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i \\ &= \int (\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i = \mathbf{X}_i\boldsymbol{\beta} = \mathbb{E}(Y_{ij}|\mathbf{b}_i = \mathbf{0}).\end{aligned}$$

Assim, existe equivalência entre a parte preditiva fixa do modelo condicional e a marginal, ou seja, a interpretação dos parâmetros $\boldsymbol{\beta}$ é a mesma em ambos modelos, a nível específico de grupo e indivíduo e a nível de população.

Entretanto, isso não ocorre no caso geral em que $g^{-1}(\cdot)$ é uma função não linear, já que a esperança da função não é a função da esperança:

$$\mathbb{E}(Y_{ij}) = \mathbb{E}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \int g^{-1}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) f_b(\mathbf{b}_i) \mathrm{d}\mathbf{b}_i \neq \mathbb{E}(Y_{ij}|\mathbf{b}_i = \mathbf{0}).$$

Para obter a esperança marginal é necessário calcular a integral, o que, em geral, não tem solução analítica.

No caso de $Y_{ij}|\mathbf{b}_i$ com distribuição de Poisson e função de ligação $\log$, têm-se

$$\log\left(\mathbb{E}(Y_{ij}|\mathbf{b}_i)\right) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i \Rightarrow g^{-1}(\mathbb{E}(Y_{ij}|\mathbf{b})) = \exp\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i\right)$$

tal que

$$\mathbb{E}(Y_{ij}) = \mathbb{E}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) \int \exp(\mathbf{z}'_{ij}\mathbf{b}_i) f_b(\mathbf{b}_i) \mathrm{d}\mathbf{b}_i \neq \exp\left(\mathbf{x}'_{ij}\boldsymbol{\beta}\right)$$

e não há correspondência entre $\boldsymbol{\beta}$ do modelo condicional e do marginal. Similarmente, não há solução analítica para a variância marginal, já que sua expressão é

$$\mathbb{V}(Y_{ij}) = \mathbb{V}(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) + \mathbb{E}(\mathbb{V}(Y_{ij}|\mathbf{b}_i)) = \mathbb{V}(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)) + \mathbb{E}(\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)).$$

Em particular, é possível obter a solução analítica, para $Y_{ij}|\mathbf{b}_i$ Poisson e ligação $\log$, quando apenas o intercepto é aleatório, isto é, $g(\boldsymbol{\mu}_i|b_{0i}) = \log\left(\boldsymbol{\mu}_i|b_{0i}\right) = \mathbf{X}_i\boldsymbol{\beta} + b_{0i}$, com $b_{0i} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2)$ (FITZMAURICE; LAIRD; WARE, 2012; FAHRMEIR et al., 2013). A solução analítica é

$$\mathbb{E}(Y_{ij}) = \exp\left(\mathbf{x}'_{ij}\boldsymbol{\beta}\right) \exp\left(0{,}5\sigma_b^2\right) = \exp\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + 0{,}5\sigma_b^2\right)$$

ou seja, os efeitos de $\mathbf{x}'_{ij}$ se mantém nos dois modelos, embora $\mathbb{E}(Y_{ij})$ depende de σ_b^2, causando um efeito no intercepto. Ainda, é possível mostrar também que

$$\mathbb{V}(Y_{ij}) = \mathbb{E}(Y_{ij}) \cdot (1 + \exp\left(1{,}5\sigma_b^2\right) - \exp\left(0{,}5\sigma_b^2\right)).$$

Como o segundo fator é maior que 1, a variância marginal é maior que a média marginal. Portanto, o modelo marginal não se encaixa na distribuição de Poisson, entretanto, considera-se Poisson com sobredispersão.

2.2.2.3 Estimação dos Parâmetros

Fahrmeir et al. (2013) aponta que, conceitualmente, inferências sobre os MLGMs possui os mesmos princípios dos MLMs sob a perspectiva da verossimilhança. No entanto, como a verossimilhança condicional em geral é não-Gaussiana e relações entre média e preditor é não linear nos MLGMs, segue que algumas partes importantes da análise não podem ser feitas analiticamente devido a falta de uma forma exata para as expressões do modelo marginal. Desta maneira, é necessário recorrer a métodos numéricos, aproximações ou abordagem Bayesiana.

Ainda, a dificuldade em determinar uma solução explícita para o modelo marginal devido a não linearidade de g^{-1} para o cenário não-Gaussiana, leva a alternativas de modelagem das variáveis discretas correlacionadas (FITZMAURICE; LAIRD; WARE, 2012). Os focos de estudos mais comuns são:

1. Modelo marginal: quando deseja-se efetuar inferência para a população, que por sua vez engloba alternativas como modelos específicos (via verossimilhança) ou alternativas que não formalizam distribuições, tais como pseudo-verossimilhança e equações de estimação generalizadas (GEE).
2. MLGM: quando deseja-se efetuar inferência para qualquer indivíduo da população.
3. MLGM com modificação no preditor linear, tal que seja marginalmente interpretável.

Nesse trabalho, o foco será apenas na abordagem do item (2).

Há inúmeras técnicas que abordam a estimação dos parâmetros $\boldsymbol{\beta}$ e os parâmetros relativos aos efeitos aleatórios $\mathbf{b}$. Entre outros autores, citamos McCulloch, Searle e Neuhaus (2011) que aborda sucintamente as técnicas mais utilizadas, tais como máxima verossimilhança, quase-verossimilhança penalizada e verossimilhança condicional e Fahrmeir et al. (2013) que retrata mais detalhadamente as técnicas da verossimilhança penalizada e processo de Monte Carlo via Cadeias de Markov, sob a perspectiva de inferência Bayesiana.

Com já apresentado no caso do modelo Gaussiano, para n grupos, a função de verossimilhança marginal é o produto das partes para os vetores observados $\mathbf{y}_i$, cada um com n_i observações, dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi) = \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\boldsymbol{\Sigma}) \mathrm{d}\mathbf{b}_i \tag{2.15}$$

Como já visto, no caso geral, a integral em (2.15) não tem solução analítica, necessitando de aproximações numéricas.

Ao trabalhar com a expansão de (2.15), considerando MLGM com apenas intercepto aleatório, é possível mostrar que

$$\begin{aligned} L(\boldsymbol{\theta}_i, \phi) &= \prod_{i=1}^{n} \int \exp \sum_{j=1}^{n_i} \left(\frac{y_{ij}\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)}{\phi^2} - c(y_{ij}, \phi) \right) \frac{\exp\left(-\mathbf{b}_i^2/2\sigma_b^2\right)}{\sqrt{2\pi\sigma_b^2}} \mathrm{d}\mathbf{b}_i \\ &= \prod_{i=1}^{n} \int h_i(b_i) \frac{\exp\left(-\mathbf{b}_i^2/2\sigma_b^2\right)}{\sqrt{2\pi\sigma_b^2}} \mathrm{d}\mathbf{b}_i \end{aligned} \tag{2.16}$$

que pode ser vista como o produto de verossimilhanças unidimensional de integrais da forma

$$\int h(b) \frac{\exp\left(-b^2/2\sigma_b^2\right)}{\sqrt{2\pi\sigma_b^2}} \mathrm{d}b$$

e quando considerado a mudança de variáveis $b = \sqrt{2\sigma_b^2}u$, tal integral se torna na forma

$$\int h(\sqrt{2\sigma_u^2}v)\frac{\exp(-v^2)}{\sqrt{\pi}}\mathrm{d}v = \int q(v)\exp(-v^2)\mathrm{d}v$$

em que $q(\cdot) = h(\sqrt{2\sigma_b^2}\cdot)/\sqrt{\pi}$.

2.2.2.3.1 Aproximação de Laplace

Uma das aproximações de integrais mais conhecidas é a aproximação de Laplace. Jiang e Nguyen (2021) descrevem o processo, para uma variável, como: suponha que o objetivo é obter uma aproximação para integrais do tipo

$$\int \exp(-q(x))\mathrm{d}x$$

em que $q(\cdot)$ é uma função bem comportada, do ponto de vista que possui valor mínimo em $x = x_0$ com $q'(x_0) = 0$ e $q''(x_0) > 0$. Então, por expansão em série de Taylor ao redor de $x = 0$, tem-se

$$q(x) = q(x_0) + \frac{1}{2}q''(x_0)(x - x_0)^2 + \cdots$$

e, consequentemente, a integral é aproximada por

$$\int \exp(-q(x))\mathrm{d}x \approx \sqrt{\frac{2\pi}{q''(x_0)}}\exp(-q(x_0)). \tag{2.17}$$

Extensões multivariadas de (2.17) são mais úteis para resolver o problema da verossimilhança. Assim, considere $q(\alpha), \alpha \in \mathbb{R}^n$, bem comportada que possui valor mínimo em $\alpha = \alpha_0$ com $q'(\alpha_0) = 0$ e $q''(\alpha_0) > 0$, onde $q'(\cdot)$ e $q''(\cdot)$ são, respectivamente, o gradiente (vetor das primeiras derivadas) e a matriz Hessiana (matriz das segundas derivadas) da função $q(\cdot)$. Então, segue que

$$\int \exp(-q(\alpha))\mathrm{d}\alpha \approx c|q''(\alpha_0)|^{-1/2}\exp(-q(\alpha_0))$$

onde c é uma constante que depende apenas da dimensão da integral e $|A|$ denota o determinante de uma matriz positiva definida A.

Em suma, a aproximação de Laplace expande o integrando em série de Taylor e é uma boa alternativa quando o número de observações por grupo/indivíduo for grande.

Com a aproximação de Laplace, é possível aplicar o método da Quase-Verossimilhança Penalizada (PQL), que consiste numa ideia semelhante ao IRWLS, isto é, lineariza μ_{ij} em torno de zero, trabalhando com $\mathbf{t}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + g(\mu_i)(y_i - \mu_i)$. O algoritmo é

1. Ajustar um MLM a partir de $\mathbf{t}_i$ e o respectivos pesos $w_{ij} = \phi\mathcal{V}^{-1}(\mu_{ij})$. Obter $\hat{\boldsymbol{\beta}}$, $\tilde{\mathbf{b}}$ e $\hat{\boldsymbol{\Sigma}}$.
2. Atualizar t_{ij} e w_{ij}, usando as estimativas obtidas no passo 1 e iterar o processo até a convergência.

Mais detalhes do processo pode ser obtido em Jiang e Nguyen (2021) e McCulloch, Searle e Neuhaus (2011), por exemplo.

Segundo Fitzmaurice, Laird e Ware (2012), o PQL tem bom desempenho computacional para variáveis quantitativas (Gamma, Normal Inversa e Poisson com parâmetro $\theta > 5$). Se a distribuição das observações for binomial, com número de ensaios pequenos, probabilidade de sucesso nos extremos ($\mu \to 0$ ou $\mu \to 1$) e com poucas medidas repetidas, as estimativas podem ser bastante enviesadas. Em suma, o desempenho do PQL depende da qualidade da aproximação Normal aos dados binomiais.

2.2.2.3.2 Quadratura Gaussiana

A Quadratura Gaussiana, também conhecida por Quadratura de Gauss-Hermite, aproxima a integral por uma soma utilizando Q pontos de quadratura igualmente espaçados e com respectivos pesos, isto é,

$$\int q(v) \exp\left(-v^2\right) \mathrm{d}v \approx \sum_{k=1}^{d} q(x_k) w_k \tag{2.18}$$

onde d é o número de quadraturas, w_k são os pesos e os pontos de avaliação x_k são designados para fornecer uma aproximação precisa no caso em $q(\cdot)$ é um polinômio. McCulloch, Searle e Neuhaus (2011) afirma que a Quadratura Gaussiana fornece a resposta exata para polinômios de graus $2d - 1$. Nos método adaptativos, o espaçamento entre os pontos é variável. Stringer e Bilodeau (2022) apresenta o processo detalhadamente.

Esse método é o mais preciso, no entanto, é computacionalmente caro. Se o número de efeitos aleatórios for pequeno, é o método mais indicado, posto que conforme a dimensão q dos efeitos aleatórios aumenta, o número de quadraturas a ser calculada cresce em uma taxa exponencial m^q (TUERLINCKX et al., 2006).

2.2.3 Inferência

Após ajustado algum modelo aos dados, um dos interesses é avaliar a precisão das estimativas, testar hipóteses sobre os parâmetros e avaliar a qualidade do ajuste em relação a outros modelos. Como a modelagem envolve duas partes, aquela que modela o preditor linear ($\mathbf{X}\boldsymbol{\beta}$) e aquela que modela a estrutura de variância e covariância $\boldsymbol{\theta}$, é necessário alguma estratégia para efetuar as análises, posto que alterações em uma das partes pode influenciar as estimativas da outra.

Entre muitos trabalhos científicos na área, tais como Hui, MÜller e Welsh (2021) e Gumedze e Dunne (2011), a estratégia é inicialmente modelar a parte aleatória, especificando a parte dos efeitos fixos da maneira mais completa possível, evitando-se que a falta de ajuste da parte fixa mascare os efeitos e/ou erros aleatórios. Após encontrar um ajuste satisfatório para a parte aleatória, buscar o modelo mais adequado para o preditor linear, isto é, a parte fixa. Para isso, comparações entre ajustes de modelos distintos são requeridas.

Tanto para MLM e MLGM, existem metodologias de inferências baseados em resultados para grandes amostras, como as estatísticas de razão de verossimilhanças (LR) e de Wald.

Dois modelos são ditos aninhados quando o mais simples é um caso particular do outro. Sendo os modelos aninhados, é possível compará-los através do teste de razão de verossimilhanças, que é válido para ambas as partes da modelagem, ou seja, tanto a parte fixa, quanto a parte aleatória.

Considere φ_k um parâmetro qualquer do modelo ajustado e deseja-se testar

$$H_0 \colon \varphi_k = \mathbf{0} \quad \text{vs} \quad H_A \colon \varphi_k \neq \mathbf{0}$$

Se φ_k for parâmetro de variância, utiliza-se $H_A \colon \varphi_k > 0$. Sob o critério da razão da verossimilhanças, ajusta-se o modelo irrestrito sob a hipótese alternativa, obtendo $\ell(\varphi_A)$ e o modelo restrito, sob H_0, obtendo $\ell(\varphi_0)$. Quando o espaço paramétrico é real, tem-se o resultado clássico, assintótico, de que a estatística, sob H_0, segue uma distribuição qui-quadrado, ou seja,

$$\Lambda = 2(\ell(\varphi_A) - \ell(\varphi_0)) \sim \chi^2_\nu$$

para $\nu = \omega_A - \omega_0$, em que ω é o número de parâmetros no modelo sob cada hipótese. No caso de φ_k ser um parâmetro de variância ($\theta_k \geqslant 0$), o valor especificado em H_0 está na fronteira de seu espaço e não há garantia do resultado assintótico.

Stram e Lee (1994) analisaram o comportamento dos componentes de variância e argumentam que que o teste Λ é conservativo, isto é, a probabilidade de significância (valor-p) calculado a partir de χ^2_ν resulta num valor maior do que o real. Os autores recomendam uma correção do valor-p, utilizando uma mistura de qui-quadrados, dada por $0.5\chi^2_{\nu-1} + 0.5\chi^2_\nu$, o que também é um resultado simplificado. No caso de MLM, Baey, CournÈde e Kuhn (2019) propuseram estimação dos pesos da mistura, o que, na prática, parece viável apenas para modelos de baixa complexidade. Existem outras propostas para a obtenção do valor-p, tais como via simulação, que são discutidas em Luke (2017) e em (PINHEIRO; BATES, 2009).

Similarmente aos parâmetros dos efeitos aleatórios, hipóteses sobre $\boldsymbol{\beta}$ também podem ser testadas através do teste de razão de verossimilhanças. A hipótese nula geral pode ser escrita como

$$H_0 \colon \mathbf{C}\boldsymbol{\beta} = \mathbf{0} \quad \text{vs} \quad H_A \colon \mathbf{C}\boldsymbol{\beta} \neq 0$$

em que $\mathbf{C}$ é uma matriz de coeficientes conhecidos das combinações lineares de $\boldsymbol{\beta}$ que se deseja testar, podendo até ser um vetor indicador de um único parâmetro em $\boldsymbol{\beta}$. A estatística do teste, sob H_0, é

$$\Lambda = 2(\ell(\boldsymbol{\beta}_A) - \ell(\boldsymbol{\beta}_0)) \sim \chi^2_\nu$$

com $\nu = \omega_A - \omega_0$ e ω é o número de parâmetros no modelo sob cada hipótese.

Intervalos de confiança para algum parâmetro também podem ser construídos pela função log-verossimilhança perfilada ($\ell_p(\beta_k)$). Para β_j em particular, faz-se a busca que maximiza $\ell_p(\beta_j)$

em uma grade de valores ao redor da estimativa obtida, encontrando a faixa que satisfaz

$$2(\ell_p(\hat{\beta}_k) - \ell_p(\beta_j)) \leqslant \chi^2_{1;\alpha},$$

em que $(1-\alpha)$ é o coeficiente de confiança e $\chi^2_{1;\alpha}$ é o respectivo quantil na distribuição qui-quadrado com 1 grau de liberdade.

Em particular para MLM, Pinheiro e Bates (2009) aponta que apenas o método ML pode ser aplicado para obter a verossimilhança, posto que REML não inclui $\boldsymbol{\beta}$ nos cálculos da verossimilhança. Além disso, Λ é anti-conservador, isto é, resulta em valor-p menor do que o real. Recomenda-se versões do teste t e F condicionais, ou seja, substituindo nas fórmulas usuais destas estatísticas, as quantidades referentes para dado $\hat{\boldsymbol{\theta}}$. Tanto o teste F quanto o teste t requerem os graus de liberdade do denominador da estatística. Tendo em vista que, sob $H_0\colon \beta_k = \beta_{k0}$, tem-se

$$T = \frac{\hat{\beta}_k - \beta_{k0}}{\widehat{\text{ep}}(\hat{\beta}_k)} \sim t_{\nu_{den}}$$

e

$$F = T^2 \sim F_{1,\nu_{den}},$$

é necessário calcular ν_{den}. A dificuldade do cálculo está no delineamento de coleta dos dados e não na ortogonalidade entre os $\hat{\beta}$'s e os efeitos dos estratos, nos quais fornecem informação para a estimação. Para efeito fixo balanceado com repeito aos efeitos aleatórios, ν_{den} é dado pelo número de graus de liberdade do estrato no qual o efeito fixo é estimado. No entanto, quando há confundimento, ou seja, um efeito fixo precisa de informação extra de um estrato para ser estimado, aproximações distintas foram propostas para seu cálculo, tais como a de Satterthwaite (1946), que utiliza o erro quadrático médio para obtenção de estimativas para os graus de liberdade e, também a de Kenward e Roger (1997), que utiliza a estatística T^2 de Hotelling para encontrar uma aproximação da estatística F.

Ainda para os MLM, os intervalos de confiança para cada β_k podem ser obtidos utilizando a distribuição t com os graus de liberdades calculados a partir das aproximações citadas anteriormente, que são chamados de tipo t-Wald, ou seja, para $(1-\alpha)$ de confiança, obtém-se

$$\left[\hat{\beta}_k \pm |t_{\nu_{den};\alpha/2}| \times \widehat{\text{ep}}(\hat{\beta}_k)\right].$$

O processo para obtenção de intervalos de confiança para θ_k é semelhante com relação aos efeitos fixos, isto é, são obtidos através da aproximação da distribuição Normal (tipo Wald) e dados por

$$\left[\hat{\theta}_k \pm |z_{\alpha/2}| \times \widehat{\text{ep}}(\hat{\theta}_k)\right].$$

Também é possível obter intervalos de confiança para parâmetros em $\boldsymbol{\theta}$ utilizando a verossimilhança perfilada.

No caso de MLGM, as outas opções de testes e construção de intervalos de confiança, são pelas estatística de Wald e função escore (TUERLINCKX et al., 2006). Por exemplo, para as

hipóteses do tipo

$$H_0\colon \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\zeta} \quad \text{vs} \quad H_1\colon \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\zeta}$$

em que $\mathbf{C}$ é uma matriz de coeficientes conhecidos de posto $k \leqslant p$, a estatística de Wald é dada por

$$w = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta})'[\mathbf{C}\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta})$$

é uma distância ponderada entre o estimador de máxima verossimilhança de $\mathbf{C}\boldsymbol{\beta}$ sob o modelo irrestrito, $\mathbf{C}\boldsymbol{\beta}$, e seu valor hipotético sobre o modelo nulo, $\boldsymbol{\zeta}$. O peso $[\mathbf{C}\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}']^{-1}$ é o inverso da matriz de variância-covariância assintótica de $\mathbf{C}\hat{\boldsymbol{\beta}}$. A matriz $\mathcal{I}(\hat{\boldsymbol{\beta}})$ é a matriz de informação de $\boldsymbol{\beta}$ que pode ser tanto a observada ou a esperada, calculadas em $\hat{\boldsymbol{\beta}}$, tal que:

$$\mathcal{I}(\hat{\boldsymbol{\beta}}) = -\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \quad \text{ou} \quad \mathcal{I}(\hat{\boldsymbol{\beta}}) = \mathbb{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\right).$$

Sob H_0, a estatística w segue, aproximadamente, a distribuição qui-quadrado com k graus de liberdade.

Por sua vez, a estatística escore, também chamada por teste do multiplicador de Lagrange, é definida por

$$u = s(\tilde{\boldsymbol{\beta}})'\mathcal{I}(\tilde{\boldsymbol{\beta}})^{-1}s(\tilde{\boldsymbol{\beta}})$$

tal que $\tilde{\boldsymbol{\beta}}$ é o vetor estimado sob o modelo restrito (H_0), e a função escore é $s(\tilde{\boldsymbol{\beta}}) = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$. A estatística u também segue, aproximadamente, a distribuição qui-quadrado com k graus de liberdade. Tanto o teste de Wald, quanto o teste por função escore, são baseados em uma aproximação quadrática para o logaritmo da função de verossimilhança através da expansão de Taylor de segunda ordem e, portanto, são aproximações para o teste de razão de verossimilhanças. Portanto, o teste da razão de verossimilhanças oferece valor-p mais acurado em relação aos demais testes, entretanto, assintoticamente, os três testes produzem o mesmo resultado (TUERLINCKX et al., 2006).

Há outros critérios de comparação de modelos, os chamados critério de informação, sendo os mais populares o Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC), definidos por

$$\begin{aligned} \text{AIC}\colon &\ -2\ell_{\text{ML}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) + 2\omega \quad \text{ou} \quad -2\ell_{\text{REML}}(\hat{\boldsymbol{\theta}}) + 2\omega \\ \text{BIC}\colon &\ -2\ell_{\text{ML}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) + 2\omega \log(N) \quad \text{ou} \quad -2\ell_{\text{REML}}(\hat{\boldsymbol{\theta}}) + \omega \log(N - p) \end{aligned}$$

para ω sendo o número de parâmetros no modelo e $N = \sum_{i=1}^{n} n_i$. O modelo com menor valor de AIC e/ou BIC é considerado o mais parcimonioso dentre aqueles sob comparação. É importante observar que os modelos comparados não necessitam ser aninhados, entretanto, quando utiliza-se o método REML, precisam conter os mesmos parâmetros de efeitos fixos. Ademais, é importante ressaltar que BIC penaliza mais a verossimilhança dos modelos com número grande de parâmetros em relação ao AIC. Também vale observar que tais critérios não informam se determinado modelo apresenta ajuste adequado.

A qualidade global do ajuste de um modelo particular também pode ser avaliada através do teste de razão de verossimilhanças para o qual o modelo alternativo é o modelo saturado. No modelo saturado, tem-se $\mathbf{E}(Y_{ij}|\boldsymbol{\beta}, \boldsymbol{\theta}_i) = y_{ij}$ para todo i e j ou seja, há tantos parâmetros quanto observações nos dados (TUERLINCKX et al., 2006). O valor da estatística LR oferecerá uma medida do quanto se perde ao usar um modelo mais simplificado.

2.2.4 Diagnósticos do Ajuste

Após ajustar qualquer modelo é importante executar técnicas de diagnóstico para avaliar a existência de evidências de violações das premissas do modelo em questão. Enquanto para o MLM a área está bem desenvolvida, com a própria natureza do modelo permitindo o cálculo de quantidades informativas ao diagnóstico, para os MLGM as possibilidades de investigação são mais limitadas. Assim, nesta seção, apresentaremos as técnicas em separado.

2.2.4.1 Técnicas para MLM

A complexidade do modelo, com suas diversas premissas, requerem a avaliação de diversos tipos de resíduos e outras medidas associadas ao ajuste do MLM.

Os resíduos marginais brutos predizem o erro marginal. Para o i-ésimo indivíduo, ele é dado por $\boldsymbol{\xi}_i = \mathbf{y}_i - \mathbb{E}(\mathbf{Y}_i) = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$ dado por:

$$\tilde{\boldsymbol{\xi}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} = \mathbf{Z}_i\tilde{\mathbf{b}}_i + \tilde{\boldsymbol{\epsilon}}_i$$

cuja variância estimada (aproximada) é

$$\widehat{\mathbb{V}(\tilde{\boldsymbol{\xi}}_i)} = \hat{\mathbf{V}}_i - \mathbf{X}_i(\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'. \tag{2.19}$$

Como no modelo linear, há a preferência pelas versões padronizadas dos resíduos para o diagnóstico do modelo. Dentro da literatura da modelagem via modelos lineares mistos, há alguns tipos padronizações dos resíduos, como, por exemplo

1. Pearson: $\tilde{\boldsymbol{\xi}}_i^P = \frac{\tilde{\boldsymbol{\xi}}_i}{\sqrt{[\hat{\mathbf{R}}_i]}}$ em que $\hat{\mathbf{R}}_i$ é a estimativa de $\mathbb{V}(\boldsymbol{\epsilon}_i)$ e $[\cdot]$ indica o respectivo elemento da sua diagonal.

2. Normalização: Para o i-ésimo indivíduo, após a decomposição de Cholesky $\hat{\mathbf{R}}_i = \hat{\boldsymbol{\Delta}}_i\hat{\boldsymbol{\Delta}}_i^{'}$

$$\tilde{\boldsymbol{\xi}}_i^N = \hat{\boldsymbol{\Delta}}_i^{-1}\tilde{\boldsymbol{\xi}}_i$$

que diferencia de Pearson apenas no caso de $\mathbf{R}$ não ser uma matriz diagonal.

Entretanto, Singer, Rocha e Nobre (2017) sugerem a padronização pela matriz de variância dos próprios resíduos marginais, dada em (2.19), e definem os resíduos marginais padronizados, para o i-ésimo grupo, como

$$\tilde{\boldsymbol{\xi}}_i^* = \hat{\boldsymbol{\Lambda}}_i^{-1}\tilde{\boldsymbol{\xi}}_i,$$

em que

$$\widehat{\mathbb{V}(\tilde{\boldsymbol{\xi}}_i)} = \hat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\Lambda}}'.$$

A principal utilidade dos resíduos marginais é a exploração de falta de ajuste na parte fixa e a presença de *outliers*. Fitzmaurice, Laird e Ware (2012) observam que o modelo marginal pode ser associado ao Modelo Linear Clássico, isto é,

$$\mathbf{Y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \boldsymbol{\epsilon}_i^* \Rightarrow \boldsymbol{\Lambda}_i^{-1}\mathbf{Y}_i = \boldsymbol{\Lambda}_i^{-1}\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Lambda}_i^{-1}\boldsymbol{\epsilon}_i$$

Assim, a verificação de tendência quando os resíduos são plotados em função das covariáveis ou dos valores ajustados é indicação de falta de ajuste.

Outra medida importante, relacionada aos resíduos marginais, foi proposta por Verbeke, Lesaffre e Brant (1998) e denominada por medida de Lesaffre Verbeke ($\mathcal{LV}$), dada por

$$\mathcal{LV}_i = \left\| \mathbf{I} - \left(\tilde{\boldsymbol{\xi}}_i^N\right)\left(\tilde{\boldsymbol{\xi}}_i^N\right)' \right\|^2$$

tal que $\|\mathbf{A}\|$ é norma de Frobenius para uma matriz $\mathbf{A}$. Com a correção proposta por Singer, Rocha e Nobre (2017), a métrica de Lesaffre Verbeke modificada fica dada por

$$\mathcal{LV}_i^* = \frac{\sqrt{\left\| \mathbf{I} - \left(\tilde{\boldsymbol{\xi}}_i^*\right)\left(\tilde{\boldsymbol{\xi}}_i^*\right)' \right\|^2}}{n_i}$$

em que n_i é o número de observações do i-ésimo indivíduo ou grupo. Uma vez que os resíduos são padronizados pela decomposição, se a estrutura em $\widehat{\mathbb{V}(\tilde{\boldsymbol{\xi}}_i^*)}$ for coerente, eles ficam próximos da independência e, portanto, os valores de $\mathcal{LV}_i^*$ são próximos de zero. Desta maneira, valores altos são indicativos de falha na modelagem da estrutura de $\mathbf{V}_i$.

Os resíduos condicionais são obtidos pela diferença entre os dados observados e os ajustados condicionais, $\tilde{\mathbf{y}}_i|\tilde{\mathbf{b}}_i = \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\tilde{\mathbf{b}}_i$, ou seja,

$$\tilde{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \tilde{\mathbf{y}}_i|\tilde{\mathbf{b}}_i = \mathbf{y}_i - (\mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\tilde{\mathbf{b}}_i)$$

cuja variância, para o vetor completo, é

$$\mathbb{V}(\tilde{\boldsymbol{\epsilon}}) = \mathbf{R}(\mathbf{V}^{-1} - \mathbf{V}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{R}$$

e a estimativa aproximada é

$$\widehat{\mathbb{V}(\boldsymbol{\epsilon})} = \hat{\mathbf{R}}\hat{\mathbf{Q}}\hat{\mathbf{R}},$$

para $\hat{\mathbf{Q}} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}$. A versão padronizada é

$$\tilde{\epsilon}_{ij}^* = \frac{\tilde{\epsilon}_{ij}}{\sqrt{\widehat{\mathbb{V}(\tilde{\boldsymbol{\epsilon}})}[ij]}}$$

na qual $\tilde{\epsilon}_{ij}$ é o resíduo condicional na j-ésima observação do i-ésimo grupo e, no denominador, o elemento da diagonal de $\widehat{\mathbb{V}(\tilde{\boldsymbol{\epsilon}})}$ na posição respectiva a esse resíduo. Os resíduos condicionais são

indicados para detectar *outliers*, homoscedasticidade e normalidade de $\boldsymbol{\epsilon}$, embora, autores como Hilden-Minton (1995) e Singer, Rocha e Nobre (2017) reconheceram que, devido ao possível confundimento entre $\tilde{\boldsymbol{\epsilon}}$ e $\tilde{b}$, tais resíduos não são apropriados para averiguação na normalidade. Eles recomendam extrair de $\tilde{\boldsymbol{\epsilon}}$ um componente minimalmente confundido de $\tilde{\mathbf{b}}$. A partir da decomposição espectral de

$$\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{Q}}\hat{\mathbf{R}}^{-1/2} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}'$$

tal que $\mathbf{L}'\mathbf{L} = \mathbf{I}_{n-p}$ e os elementos na diagonal de $\boldsymbol{\Lambda}$ são os autovalores $0 < \lambda_{n-p} \leqslant \cdots \leqslant \lambda_2 \leqslant \lambda_1 \leqslant 1$, os resíduos minimamente confundidos padronizados são

$$\tilde{\boldsymbol{\epsilon}}^{\dagger} = \frac{\mathbf{C}\tilde{\boldsymbol{\epsilon}}}{\sqrt{\text{diag}(\mathbf{C}\hat{\mathbf{R}}\hat{\mathbf{Q}}\hat{\mathbf{R}}\hat{\mathbf{C}}')}}$$

tal que $\mathbf{c}_k'$, a k-ésima linha de $\mathbf{C}$, é $\mathbf{c}_k' = \lambda_k^{-1/2}\ell_k'$ e ℓ_k' é a k-ésima coluna de $\mathbf{L}$. Pela decomposição, apenas $n-p$ resíduos são extraídos, aquelas combinações de $\tilde{\boldsymbol{\epsilon}}$ independentes e são consideradas apropriadas para checar a normalidade de $\boldsymbol{\epsilon}$.

Por fim, os resíduos dos efeitos aleatórios são as predições dos efeitos aleatórios dadas por $\tilde{\mathbf{b}}$. Sob normalidade de $\mathbf{b}$ e a suposição de não confundimento entre os efeitos, a forma quadrática, representando a distância de Mahalanobis entre $\tilde{\mathbf{b}}_i$ e seu valor esperado $\mathbf{0}$, é dada por

$$\mathcal{M}_i = \tilde{\mathbf{b}}_i' \left(\mathbb{V}(\tilde{\mathbf{b}}_i - \mathbf{b}_i)\right)^{-1} \tilde{\mathbf{b}}_i$$

e segue a distribuição χ^2_q em que q é a dimensão de $\mathbf{b}_i$. O gráfico QQ-plot qui-quadrado de $\mathcal{M}_i$ é utilizado para checar a normalidade de $\mathbf{b}_i$.

Em suma, a partir de Singer, Rocha e Nobre (2017), apresentamos a Tabela 1 para resumir as ferramentas e auxiliar o diagnóstico.

Tabela 1 – Uso das estatísticas específicas para diagnóstico de ajuste de um Modelo Linear Misto.

Diagnóstico	Tipo de Resíduo	Gráfico	Procurar
Falta de ajuste ($\mathbf{X}\boldsymbol{\beta}$)	Marginal	$\tilde{\boldsymbol{\xi}}^*$ vs $\mathbf{X}$ e $\hat{\mathbf{y}}$	Tendência
Outlier em y_{ij}	Marginal	Índice de $\tilde{\boldsymbol{\xi}}^*$	Pontos destaques
Covariância intra-grupo ($\mathbf{V}_i$)	Marginal	Índice de $\mathcal{LV}_i^*$	Pontos destaques
Outlier em y_{ij}	Condicional	Índice de $\tilde{\boldsymbol{\epsilon}}^*$	Pontos destaques
Homoscedasticidade de ϵ_{ij}	Condicional	$\tilde{\boldsymbol{\epsilon}}^*$ vs $\hat{\mathbf{y}}$ e $\mathbf{X}$	Dispersão dos pontos
Normalidade de ϵ_{ij}	Condicional	QQ-Norm de $\tilde{\boldsymbol{\epsilon}}^{\dagger}$	Fuga da reta
Outlier nos grupos ($\mathbf{b}_i$)	Efeito aleatório	Índice para $\mathcal{M}_i$	Pontos destaque
Normalidade de $\mathbf{b}_i$	Efeito aleatório	QQ-χ^2_q de $\mathcal{M}_i$	Fuga do envelope

2.2.4.2 Técnicas para MLGM

O ferramental para diagnóstico de ajuste de Modelos Mistos Generalizados é um tanto quanto restrito. Entretanto, as técnicas que diversos autores utilizam são, em grande parte, as mesmas para Modelos Lineares Generalizados. Com isso em mente, para os diagnósticos de

ajuste dos modelos mistos generalizados presentes neste trabalho, serão utilizados os seguintes recursos gráficos

- Gráfico Seminormal: Útil para diagnosticar sobredispersão. Verificar se há fuga de envelope;
- Gráfico Quantil-Quantil: Útil para diagnosticar a normalidade dos resíduos. Verificar se há fuga da reta;
- Resíduos de Pearson *versus* Valores Ajustados: Útil para diagnosticar a linearidade do modelo. Procura-se por padrões sistemáticos dos resíduos;
- Resíduos de trabalho *versus* Preditor Linear: Útil para diagnosticar se a função de ligação está adequada para o modelo. Procura-se por resíduos altos e/ou padrões sistemáticos.

É importante ressaltar que os resíduos do tipo trabalho, segundo Dunn e Smyth (2018), são definidos por

$$e_i = z_i - \hat{\eta}_i,$$

tal que $z_i = \eta_i + \frac{\mathrm{d}\eta_i}{\mathrm{d}\mu_i}(y_i - \mu_i)$.

2.2.5 Predições para Novas Observações

Após o diagnóstico e seleção de um modelo útil para o problema, pode ser do interesse do pesquisador efetuar predições para novas observações. Há duas propostas que podem ser aplicáveis, dependendo da natureza e complexidade do modelo, usando as propriedades teóricas dos estimadores, induzidas pelos processos de estimação dos modelos, ou por simulação, esta última inspirada na abordagem Bayesiana, já que a abordagem frequencista está relacionada à Bayesiana sob distribuições a priori não informativas (GELMAN; HILL, 2007). A alternativa via simulação é conveniente para o MLGM pois permite a incorporação de todas as fontes de incertezas no processo.

Em particular, considerando um modelo com apenas efeito aleatório de intercepto, para um dado indivíduo i podemos precisar predizer a resposta $y_i(\mathbf{x}_*)$ para $\mathbf{x}_*$ representando uma condição para a qual não há dados. Usando a distribuição condicional de $Y|b$, temos que o preditor pontual de $Y_i(\mathbf{x}_*)$ é

$$\tilde{Y}_i(\mathbf{x}_*)|\tilde{b}_i = \mathbf{x}_*'\hat{\boldsymbol{\beta}} + \tilde{b}_i + \epsilon(\mathbf{x}_*) = \mathbf{x}_*'\hat{\boldsymbol{\beta}} + \tilde{b}_i,$$

com variância aproximada por

$$\mathbb{V}(\tilde{Y}_i(\mathbf{x}_*)|\tilde{b}_i) = \mathbf{x}_*'\mathbb{V}(\hat{\boldsymbol{\beta}})\mathbf{x}_* + \mathbb{V}(\epsilon(\mathbf{x}_*))$$

cuja estimativa, para o modelo homocedástico, se reduz a $\hat{\mathbb{V}}(\tilde{y}_i(\mathbf{x}_*)|\tilde{b}_i) = \mathbf{x}_*'\hat{\mathbb{V}}(\hat{\boldsymbol{\beta}})\mathbf{x}_* + \hat{\sigma}^2$, permitindo a obtenção de intervalo de predição usando a teoria usual sob normalidade.

Para modelos mais complexos, a simulação a partir da distribuição a posteriori é uma via escapatória da dificuldade numérica em métodos de aproximação, principalmente no cálculo da precisão da predição. Note que, na aproximação apresentada acima, ignorou-se a incerteza na predição de $\tilde{b}_i$.

Assim, quantificar as incertezas sobre a predição através de simulação é um caminho alternativo. Um esquema, esboçado por Gelman e Hill (2007), página 273, serve como ponto de partida para modelos mais complexos. Os passos são:

1. Estimar os coeficientes $\boldsymbol{\beta}$, b_i e σ, utilizando os dados disponíveis e obtenha $\hat{y}(\mathbf{x}_*)$ pela equação do preditor linear ($\hat{y}(\mathbf{x}_*) = \mathbf{x}'_*\hat{\boldsymbol{\beta}} + \tilde{b}_i$);

2. Simular n_{sims} realizações aleatórias para $y(\mathbf{x}_*)$ a partir de uma distribuição gaussiana com média $\hat{y}(\mathbf{x}_*)$ e variância $\hat{\sigma}^2$.

Tais simulações são centradas em $\mathbb{E}(Y(\mathbf{x}_*))$ com variabilidade σ^2, representando a incerteza do erro aleatório associado à futura observação e podem ser utilizadas para obtenção de intervalos de predição. Vale notar que esse esquema de simulação é bem simplificado pois ignora as incertezas associadas ao processo de estimação dos parâmetros, o que pode ser implementado facilmente.

A ideia da simulação pode ser estendida aos MLG, MLM e MLGM. Para o modelos mistos, as predições podem ser para novas observações em grupos existentes no banco de dados ou para novas observações em novos grupos, aumentando a dificuldade da precisão preditiva analítica do modelo (GELMAN; HILL, 2007). Tuerlinckx et al. (2006) afirma que predições para MLGM a partir de abordagem Bayesiana, a qual corresponde à Inferência Clássica, quando a informação a priori é não informativa, são computacionalmente mais eficientes.

Seguindo os mesmos princípios para o MLM, as predições, em MLGM, para novas observações em grupos já existentes, são baseadas nas estimativas dos parâmetros do modelo, incluindo os efeitos fixos $\boldsymbol{\beta}$ e os efeitos aleatórios $\mathbf{b}$, e aplicando-se a transformação inversa da função de ligação, ou seja

$$\tilde{Y}_i(\mathbf{x}_*)|\tilde{\mathbf{b}}_i = h\left(\mathbf{x}'_*\hat{\boldsymbol{\beta}} + \mathbf{z}'_*\tilde{\mathbf{b}}\right)$$

em que $\hat{\boldsymbol{\beta}}$ são as estimativas dos efeitos fixos estimados, $\tilde{\mathbf{b}}$, são as predições dos efeitos aleatórios e $h(\cdot) = g^{-1}(\cdot)$ a inversa da função de ligação. O processo da simulação preditiva, segundo Gelman e Hill (2007), é dado por

1. Gerar n_{sims} amostras dos coeficientes de efeitos fixos e efeitos aleatórios de acordo com suas distribuições a posteriori associadas, ou seja,

$$\hat{\boldsymbol{\beta}}_k \sim \mathcal{N}_p(\hat{\boldsymbol{\beta}};\ \hat{\mathbb{V}}(\hat{\boldsymbol{\beta}})) \text{ e} \tag{2.20}$$

$$\tilde{\mathbf{b}}_k \sim \mathcal{N}_q(\tilde{\mathbf{b}};\ \boldsymbol{\Delta}[\boldsymbol{\Delta}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\Delta} + \mathbf{I}]^{-1}\boldsymbol{\Delta}), \tag{2.21}$$

   em que $\boldsymbol{\Delta}$ foi definido em (2.3).

2. Para cada amostra de parâmetros $(\hat{\boldsymbol{\beta}}_k, \tilde{\mathbf{b}}_k)$ simulada, calcula-se uma predição simulada dada por

$$\tilde{\mathbb{E}}(Y_{k_*}|\tilde{\mathbf{b}}_k) = h(\mathbf{x}'_*\hat{\boldsymbol{\beta}}_k + \mathbf{z}'_*\tilde{\mathbf{b}}_k),$$

resultando numa distribuição preditiva para cada observação, que reflete a incerteza associada à estimação dos parâmetros do modelo.

Esse processo geral vale tanto para MLM quanto para MLGM, com a modificação de que para o MLM, são gerados n_{sims} amostras de $\sigma^2 = \hat{\sigma}^2\sqrt{(n-p)/\mathcal{X}}$, em que $\mathcal{X}$ são valores gerados a partir da distribuição χ^2 com $n-p$ graus de liberdade, sendo p o número de parâmetros de efeitos fixos do modelo, que posteriormente são substituídos na expressão de $\mathbb{V}(\hat{\beta})$ em (2.20).

Um passo a mais para incluir a variabilidade aleatória é, usando $\tilde{\mathbb{E}}(Y_{k_*}|\tilde{\mathbf{b}})$, simular uma realização da variável aleatória a partir do modelo probabilístico associado à variável resposta.

No caso de predição de observações para novos grupos, por exemplo, o grupo i_*, seguem-se os mesmos passos anteriores, exceto que o efeito aleatório do novo grupo é simulado via a expressão

$$\tilde{\mathbf{b}}_{i_*k} \sim \mathcal{N}_q(\mathbf{0};\ \hat{\boldsymbol{\Sigma}}),$$

ao invés da equação (2.21).

3 Material e Metodologia

3.1 Metodologia

A ocorrência de eventos adversos severos como desastres ambientais, guerras e pandemias afeta a saúde e a sobrevivência da população, seja humana ou de animais. No caso particular da pandemia de COVID-19, a mortalidade na população brasileira foi uma das mais altas do mundo. No entanto, uma pandemia da magnitude e características experienciada pode ter tido efeitos diretos e indiretos na saúde das pessoas de forma que a apenas a contagem de óbitos causados pela doença específica não informa claramente sobre os impactos sofridos.

WHO (2023) considera que a quantificação do excesso de mortalidade ocorrido no período é essencial para a compreensão do preço pago pela população, em termos de vidas perdidas, não só pela doença em si, como devido às restrições de locomoção, de trabalho, de subsistência e demais dificuldades sofridas pela sobrecarga dos serviços de saúde, assim como pelo receio de buscar tratamentos por outras doenças.

O excesso de mortalidade durante um evento adverso é definido como sendo a diferença entre a mortalidade observada e a mortalidade que seria esperada, sob situação de normalidade, ou seja, supondo que o evento não ocorreu.

Considerando y_{ij} o número de mortes observado na j-ésima semana do i-ésimo ano e E_{ij} a previsão esperada do número de mortes na j-ésima semana do i-ésimo ano, na hipótese da pandemia não ter acontecido, o excesso de mortes δ_{ij} na j-ésima semana do i-ésimo ano é definido, matematicamente, por

$$\delta_{ij} = y_{ij} - E_{ij}.$$

Com isso, a grande dificuldade do cálculo do excesso de mortes está na predição do número de mortes esperado, tendo em vista que os dados sobre o número de mortes são coletados pelo Ministério da Saúde, Governo Brasileiro, desde 1979. Diversas metodologias foram propostas na bibliografia, desde médias das contagens em um grupo de anos prévios (dados históricos) até modelos bastante sofisticados. Os modelos mistos e modelos de séries temporais tais como aqueles aplicados em Verbeeck et al. (2021), Santos et al. (2021), Palacio-MejÍa et al. (2022) e Colonia et al. (2023) são possibilidades de complexidade moderada. Neste trabalho, utilizaremos modelos mistos seja de caráter linear ou linear generalizado para o ajuste aos dados de mortalidade

ao longo do tempo, isto é, dos cinco anos anteriores à pandemia mais as 9 primeiras semanas do ano de 2020, totalizando mortalidade acumulada semanalmente para 269 semanas.

Retomando as notações das seções 2.2.1 e 2.2.2, e seguindo a formulação usual para modelar comportamentos periódicos em epidemiologia, como é o caso da mortalidade brasileira ao longo do ano, utilizamos o primeiro termo da série de Fourier. Desse modo, a expressão básica do preditor linear do modelo é dada por

$$g(\mathbb{E}(Y_{ij}|\mathbf{b}_i)) = \eta_{ij} = (\beta_0+b_{0i})+(\beta_1+b_{1i})\sin\left(\frac{2\pi}{52}x_j\right)+(\beta_2+b_{2i})\cos\left(\frac{2\pi}{52}x_j\right)+\beta_3 t_{ij}+\log \text{pop}_{ij} \tag{3.1}$$

em que $g(\cdot)$ é a função de ligação, Y_{ij} representa o número de mortes na semana j do ano i ($j = 1, \ldots, 52$ e $i = 1, \ldots, 6$), $x_j = j$, β's são os efeitos fixos e $\mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})' \sim \mathcal{N}_3(\mathbf{0}, \mathbf{\Sigma})$ são os efeitos aleatórios. Note que, além dos termos da série de Fourier, foi acrescentado um termo linear de tempo tal que $t_{ij} \in \left\{\frac{1}{269}, \frac{2}{269}, \ldots, 1\right\}$ é o tempo, em semanas de 1 a 269, em escala codificada.

Os modelos utilizados para a estimação das mortes esperadas em 2020, sob a hipótese da não ocorrência da pandemia, foram Normal com ligação identidade, Poisson e Binomial Negativa, ambas com ligação logarítmica. É importante ressaltar que a utilização da população, uma variável *offset*, não é necessária para o modelo Gaussiano, visto que ela é utilizada para representar taxas de exposição das observações em modelos de Poisson e Binomial Negativa.

Para os ajustes dos modelos estatísticos no presente trabalho, foi utilizado o software R (R Core Team, 2022) de código aberto. Os pacotes `nlme` (PINHEIRO; BATES; R Core Team, 2023) e `lme4` (BATES et al., 2015) foram utilizados para ajustar modelos lineares mistos e lineares generalizados mistos, respectivamente. Embora o último possa ajustar ambos os tipos de modelos, o primeiro é mais conveniente para o caso linear, para a aplicação das técnicas de análise de diagnóstico. Já para diagnósticos, foram utilizados os pacotes `hnp` (MORAL; HINDE; DEMÉTRIO, 2017) e a função `residdiag.lmm` de autoria de Singer, Rocha e Nobre (2017). Por fim, para gráficos, foi utilizado `ggplot2` (WICKHAM, 2016).

Uma vez estimados os parâmetros dos modelos, as predições para mortalidade para o período pandêmico de 2020 e 2021, que é caracterizado pelas semanas 10 a 52 do ano de 2020 e pelo ano de 2021, foram obtidas. A incorporação de todas as fontes de incerteza no processo preditivo, cometidas pelos modelos, foi realizado via processo de simulação. Para isso foi construída uma função que se apropria dos objetos de saída da função `sim` do pacote `arm` (GELMAN; SU, 2024) adicionando erros às estimativas dos parâmetros e a variabilidade aleatória na predição, tornando-se uma adaptação da função `predict.sim.merMod`, também do pacote `arm`.

3.2 Material

Para estudar e, consequentemente, gerar pesquisas sobre a mortalidade no Brasil, o Ministério da Saúde criou o Sistema de Informação sobre Mortalidade, SIM (Ministério da Saúde, 2023c), em 1975, implementando mais de quarenta modelos de instrumentos para regularizar a captura dos dados sobre a mortalidade no país. Entretanto, somente em 1979 o SIM foi informatizado e, após doze anos, teve a coleta dos dados distinguida por Estados e Municípios, através das Secretarias de Saúde, tornando-o uma ferramenta imprescindível para a gestão no setor de saúde que subsidia tomadas de decisão em inúmeras áreas da assistência de saúde (Ministério da Saúde, 2023a).

O documento básico vital que alimenta o banco de dados do SIM é o Documento de Óbito (DO), que é declarado por médicos, impresso e assinado em três vias pré-numeradas sequencialmente, conforme prevê o Artigo 115 do Código de Ética Médica, Artigo 1º da Resolução nº 1779/2005 do Conselho Federal de Medicina e a Portaria SVS nº 116/2009 e deve ser enviado ao Cartório de Registro Civil para liberação do sepultamento, assim como as demais medidas comumente tomadas em relação ao falecimento de uma pessoa.

Conforme a Portaria SVS nº 116/2009, de 11/02/2009, da Secretaria de Vigilância em Saúde, as Declarações de Óbitos são preenchidas por unidades notificantes de óbito e recolhidas pelas Secretarias Municipais de Saúde, donde são digitadas, processadas e implementadas no SIM, permitindo a formulação de indicadores de mortalidade por causas específicas de interesse à análise e avaliação dos sistemas locais, microrregionais, estaduais e nacional de saúde, permitindo a comparação epidemiológica do Brasil com as de outros países a partir da mortalidade, que pode ser estratificada por variáveis tais como causas primárias e secundárias de óbito, sexo, idade, município e estado de ocorrência, estado civil e escolaridade, ou seja, por meio de fatores de mortalidade e sociodemográficas.

Para o presente estudo, foram selecionados no SIM, dados sobre a mortalidade brasileira durante os anos de 2015 até 9ª semana de 2020 para realização do estudo comportamental brasileiro e a partir da 10ª semana de 2020 até 2021 para realização do estudo de excesso de mortalidade. Os dados originais foram pré-processados e agrupados por semana epidemiológica (definição US CDC) de cada ano. Ademais, os dados sobre a projeção da população, utilizados como variável *offset* na modelagem de Poisson e Binomial Negativa, tem como fonte original o IBGE, Instituto Brasileiro de Geografia e Estatística (2024), mas já se encontram pré tratados e disponíveis via o trabalho de Santos et al. (2021).

4 Resultados

Nesse capítulo, apresentamos, na seção 4.1, os resultados dos ajustes, alguns diagnósticos e comparações entre os modelos considerados como candidatos para predizer a mortalidade brasileira no ano 2020, caso a pandemia não tivesse ocorrido. Na seção 4.2 exploramos a precisão das predições e, por fim, os resultados do excesso de mortalidade são apresentados.

4.1 Ajustes dos Modelos

O perfil de mortalidade no Brasil ao longo do tempo, no período da 10^a a 52^a semanas, exibe um padrão cíclico anual nítido, com aumento do número de óbitos nas semanas de inverno e queda nos meses mais quentes, o que pode ser visto pelos pontos na Figura 1. Além desse comportamento periódico, nota-se também uma tendência linear crescente ao longo do tempo. Tal padrão não é particular da mortalidade brasileira, dados de diversos países mostram efeito sazonal de mortes. Com base nesse perfil e também com suporte em inúmeros trabalhos que

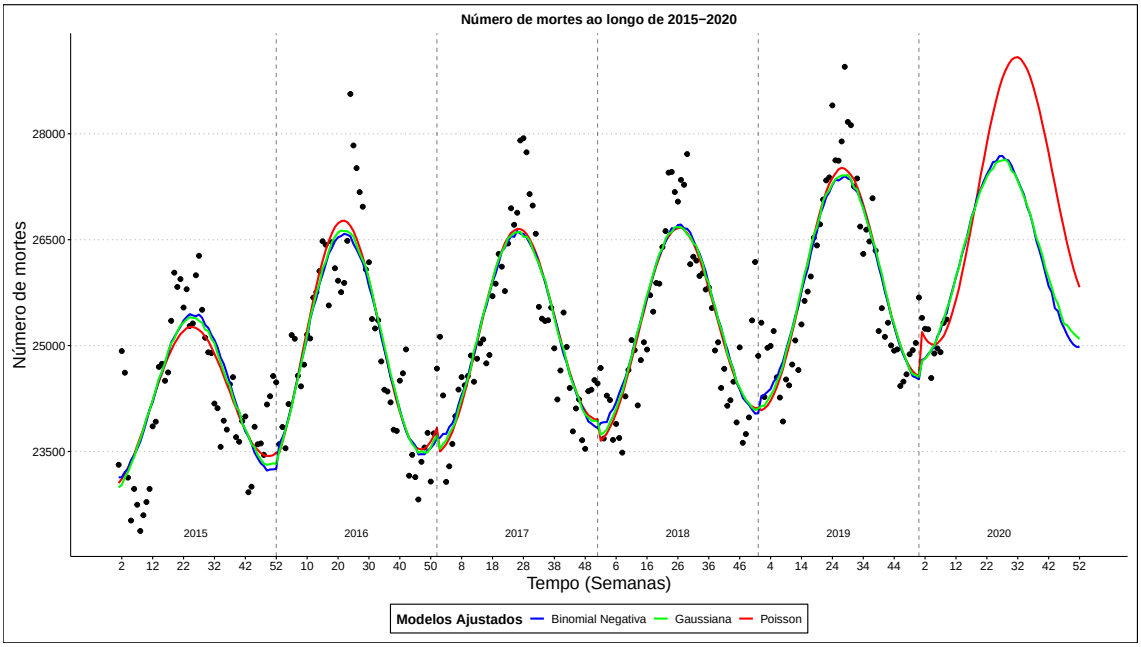


Figura 1 – Comportamento periódico da mortalidade brasileira a partir da 1^a semana de 2015 até a 9^a semana de 2020, juntamente com os modelos mistos ajustados.

modelaram mortalidade, a parte preditiva dos modelos ajustados incluiu efeito linear de tempo e as transformações pelo primeiro termo da série de Fourier. No arcabouço dos modelos de efeitos mistos, em princípio, os coeficientes associados ao tempo são parâmetros fixos e cada ano é visto como um grupo de efeito aleatório. Tal efeito aleatório (intercepto) se responsabiliza pela modelagem da heterogeneidade, entre os anos, da posição vertical das curvas. Efeitos aleatórios associados aos outros termos do modelo se responsabilizam pela modelagem da heterogeneidade da forma das curvas ao longo dos anos. Essa estrutura foi utilizada nos três tipos de modelos considerados neste trabalho.

4.1.1 Modelo Gaussiano

Para o modelo Gaussiano misto, cuja função de ligação é identidade, a expressão da mortalidade condicional é dada por

$$Y_{ij}|\mathbf{b}_i = \eta_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \operatorname{sin}\left(\frac{2\pi}{52}x_j\right) + (\beta_2 + b_{2i}) \operatorname{cos}\left(\frac{2\pi}{52}x_j\right) + \beta_3 t_{ij} + \epsilon_{ij},$$

cuja notação e premissas estão explicitadas ao redor da equação (3.1), exceto a adição do termo de erro aleatório com $\boldsymbol{\epsilon} \sim \mathcal{N}_{269}(\mathbf{0};\ \sigma^2\mathbf{I})$. Em princípio, ajustou-se o Modelo 1 com $\omega = 11$ parâmetros, 4 fixos, 6 relacionados à $\boldsymbol{\Sigma}$, a matriz de variância-covariância dos termos b_{0i}, b_{1i} e b_{2i}, simétrica e positiva definida, ou seja, uma variância para cada termo aleatório ($\sigma_0^2,\ \sigma_1^2,\ \sigma_2^2$) e as covariâncias ($\sigma_{01}, \sigma_{02}$ e σ_{12}) e, por fim, o parâmetro σ^2. O Modelo 2, com $\omega = 8$ parâmetros, buscou a simplificação do Modelo 1 e considerou $\boldsymbol{\Sigma}$ sendo uma matriz diagonal, isto é, sem correlação entre os efeitos aleatórios.

As estatísticas globais dos ajustes destes dois modelos estão expostos na Tabela 2. Os valores dos critérios de informação são um pouco menores para o modelo mais simples e valor-p não revela evidência de falta de ajuste do modelo mais simples em relação ao Modelo 1. Assim, o Modelo 2 é considerado mais parcimonioso por conter número menor de parâmetros.

Ao realizar o diagnóstico dos Modelos 1 e 2, de acordo com a Tabela 1, utilizando o função `residdiag.nlme` (SINGER; ROCHA; NOBRE, 2017), foi observado padrões semelhantes nos gráficos. É possível reparar que na Figura 2, nos quais, para análise de violação de premissas, busca-se se há alguma tendência nos resíduos. Nota-se uma tendência de valores positivos altos, tanto para valores ajustados baixos quanto para altos, indicando certo caráter tendencioso parabólico. A explicação, com certo suporte na Figura 1 é que o modelo periódico não aproxima bem nos picos, tanto superiores quanto inferiores. Com isso, há certa falta de ajuste, porém

Tabela 2 – Estatísticas globais da qualidade de ajuste dos modelos Gaussinos

Modelo	ω	AIC	BIC	logLik	LR	valor-p
1	11	4226,11	4265,59	−2102,05		
2	8	4224,25	4252,89	−2104,13	4,145	0,246

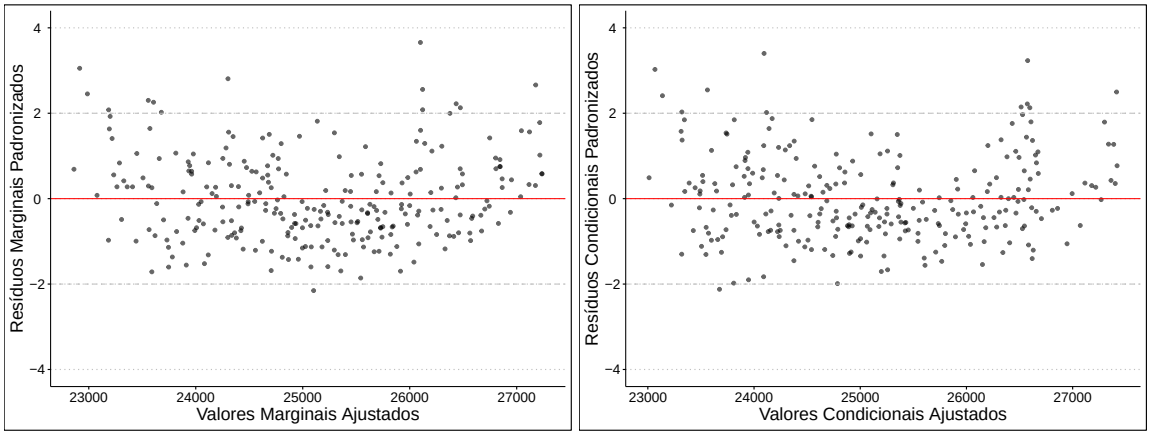


Figura 2 – Valores marginais ajustados *versus* Resíduos marginais padronizados (esquerda) e Valores condicionais ajustados *versus* Resíduos condicionais padronizados (direita) para o Modelo 2.

consideramos que, no geral, a aproximação é útil. Entretanto, nada muito suspeito é revelado contra a homoscedasticidade da variabilidade em função dos valores ajustados.

Na Figura 3, com relação ao Índice de Lesaffre e Verbeke, é possível reparar que o ponto 1, referente ao ano de 2015, é destacado, com isso, indicando leve falta de ajuste para correlação intra-grupo. Além disso, não se detecta violação quanto a normalidade dos efeitos aleatórios, posto que não há fuga do envelope no QQ-χ^2_q da distância de Mahalanobis. Por fim, na Figura 4, não indica violação na suposição de normalidade dos erros, pelo fato dos pontos estarem próximo da reta e sem fuga do interior do envelope.

A partir das estatísticas globais dos modelos ajustados (Tabela 1) e, também, dos gráficos de diagnósticos, verifica-se que o ajuste do Modelo 2 está relativamente aceitável, podendo ser considerado como sendo o mais parcimonioso. Na sequência, este modelo será denotado por Modelo Normal Misto. A Tabela 3 exibe as informações sobre as estimativas de seus parâmetros.

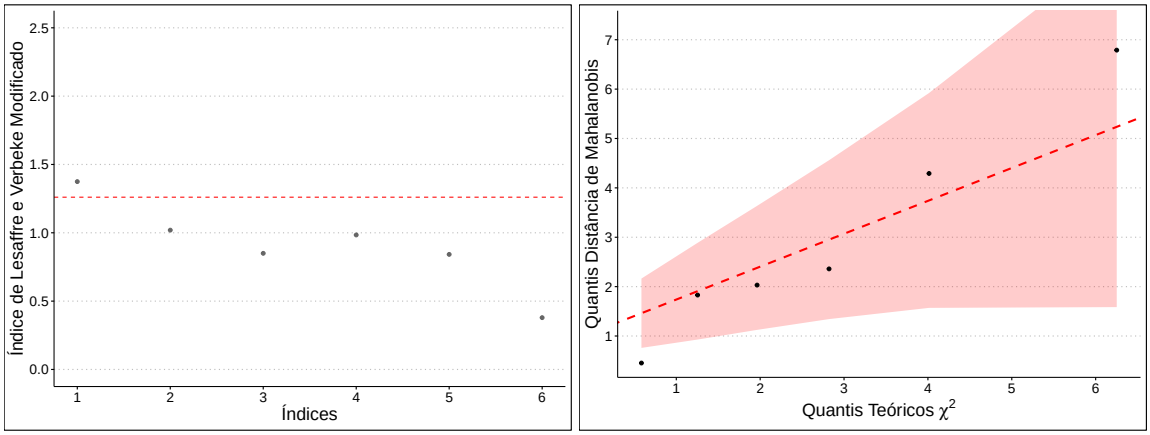


Figura 3 – Índice de Lesaffre e Verbeke (esquerda) e QQ-χ^2_q da distância de Mahalanobis, $\mathcal{M}_i$ (direita) para o Modelo 2.

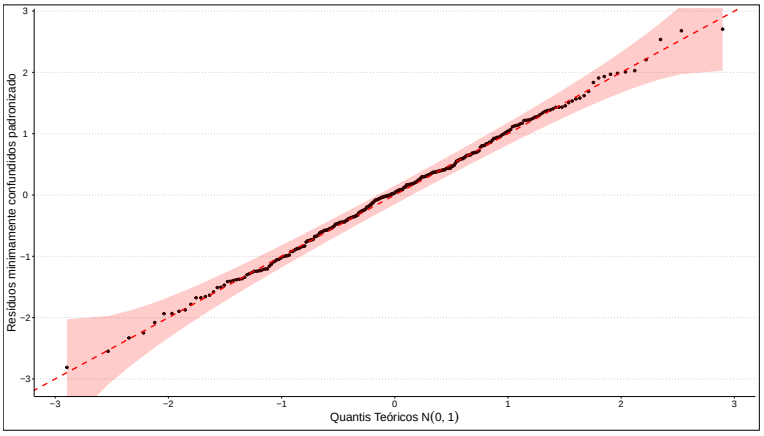


Figura 4 – QQ-Normal dos resíduos minimamente confundidos padronizados para o Modelo 2.

Embora o termo relativo à transformação seno, na parte fixa, não foi significante, ele não será removido já que a variabilidade estimada de seu componente aleatório se mostra importante ($\hat{\sigma}_1 = 405{,}23$). Outra justificativa é que o primeiro termo da série de Fourier inclui os dois componentes, seno e cosseno. As estimativas dos efeitos fixos β_1 e β_2, uma positiva e a outra negativa, representam os efeitos dos componentes para capturar o padrão periódico ods dados. Além do padrão periódico, a estimativa de β_3 indica incremento linear da altura da curva, resultando em aumento da mortalidade ao longo do tempo, conforme esperado, já que o tamanho da população também aumenta. As estimativas dos componentes de variância indicam alta heterogeneidade entre os anos, tanto com respeito ao intercepto quanto aos termos responsáveis pela periodicidade. Mas a maior variabilidade é a nível de observação, com desvio padrão de $\hat{\sigma} = 629{,}70$. O perfil da curva ajustada é exibida pela linha verde na Figura 1.

Tabela 3 – Estimativas dos parâmetros do Modelo Normal Misto e respectivas significâncias para os efeitos fixos

Efeitos Fixos				
Parâmetro	Estimativa	Erro Padrão	Valor t	Valor p
β_0	24195,62	185,90	130,15	≈ 0
β_1	138,23	181,62	0,76	0,45
β_2	$-1367{,}60$	98,06	$-13{,}95$	≈ 0
β_3	1923,53	311,55	6,17	≈ 0
Efeitos Aleatórios				
	σ_0	σ_1	σ_2	σ
Estimativa	199,06	405,23	186,48	629,70

4.1.2 Modelo de Poisson

Para o modelo de Poisson, inclui-se o logaritmo da projeção do tamanho população, em cada semana, como variável *offset*, conforme indicado na equação (3.1), ou seja, o preditor linear,

Tabela 4 – Estatísticas globais da qualidade de ajuste dos modelos Poisson

Modelo	ω	AIC	BIC	logLik	Deviance	χ^2	$\neq$GL	Pr$(>\chi^2)$
3	6	7353,7	7375,2	$-3670{,}8$	7341,7			
2	7	7212,5	7237,6	$-3599{,}2$	7198,5	143,181	1	≈ 0
1	10	7211,6	7247,6	$-3595{,}8$	7191,6	6,845	3	0,077

com a função de ligação logarítmica:

$$\log\left(\mathbb{E}(Y_{ij}|\mathbf{b}_i)\right) = (\beta_0+b_{0i})+(\beta_1+b_{1i})\sin\left(\frac{2\pi}{52}x_j\right)+(\beta_2+b_{2i})\cos\left(\frac{2\pi}{52}x_j\right)+\beta_3 t_{ij}+\log \text{pop}_{ij}.$$

Uma observação a ser destacada é que o IBGE faz projeção apenas anual e, então, os valores dessa variável é constante ao longo de cada ano.

Foram realizados três ajustes em prol de encontrar o modelo mais parcimonioso a partir das técnicas de diagnósticos simples disponíveis, tal como verificação dos resíduos e dos critérios informativos. Os dois primeiros ajustes foram feitos seguindo a mesma lógica dos modelos gaussianos, isto é, removendo a covariância entre os efeitos aleatórios, permitindo que a matriz $\mathbf{\Sigma}$ de variância-covariância seja diagonal ou uma simétrica positiva definida qualquer. Os resultados comparativos estão na Tabela 4. O Modelo 2, com matriz de variância-covariância diagonal, comparado ao Modelo 1, apresentou valores dos critérios similares. A probabilidade de significância, pela estatística qui-quadrado, embora menor que 0,10, não indica evidência contra o Modelo 2 e, prezando pela parcimônia, ele será considerado para análise mais detalhada.

Quando os efeitos aleatórios do Modelo 2 são inspecionados (Tabela 5), verifica-se que a estimativa da variância do efeito aleatório do componente do cosseno é muito pequena, levando à possibilidade de simplificação do modelo. Pensando nisso, o Modelo 3 foi ajustado, removendo tal efeito aleatório para verificar se há melhora. As estatísticas globais do ajuste também estão presente na Tabela 4.

Observa-se que tanto os valores dos critérios de informação de Akaike e Bayesiano,

Tabela 5 – Estimativas dos parâmetros do Modelo Poisson Misto, e respectivas significâncias para os efeitos fixos

	Efeitos Fixos			
Parâmetro	Estimativa	Erro Padrão	Valor z	Pr$(>\|z\|)$
β_0	$-9{,}0723$	0,0104	$-870{,}859$	≈ 0
β_1	0,0004	0,0098	0,043	0,966
β_2	$-0{,}0553$	0,0032	$-17{,}567$	≈ 0
β_3	0,1043	0.0119	8,737	≈ 0

	Parâmetros de Variabilidade		
	σ_0	σ_1	σ_2
Estimativa	$3{,}736\times10^{-4}$	$5{,}576\times10^{-4}$	$4{,}915\times10^{-5}$

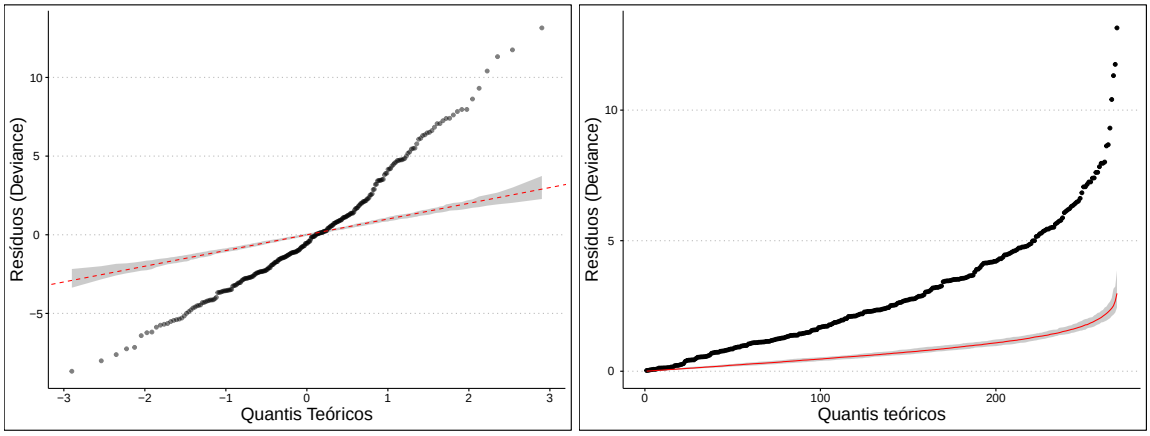


Figura 5 – Gráfico QQ-Normal dos resíduos tipo Deviance (esquerda) e Gráfico seminormal dos resíduos tipo Deviance (direita) para o Modelo Poisson Misto.

quanto o teste Qui-Quadrado, indicam que ao remover o componente aleatório do cosseno, há piora significativa no modelo. Portanto, o modelo mais parcimonioso entre os ajustados foi o segundo, que será denotado, na sequência por Modelo Poisson Misto.

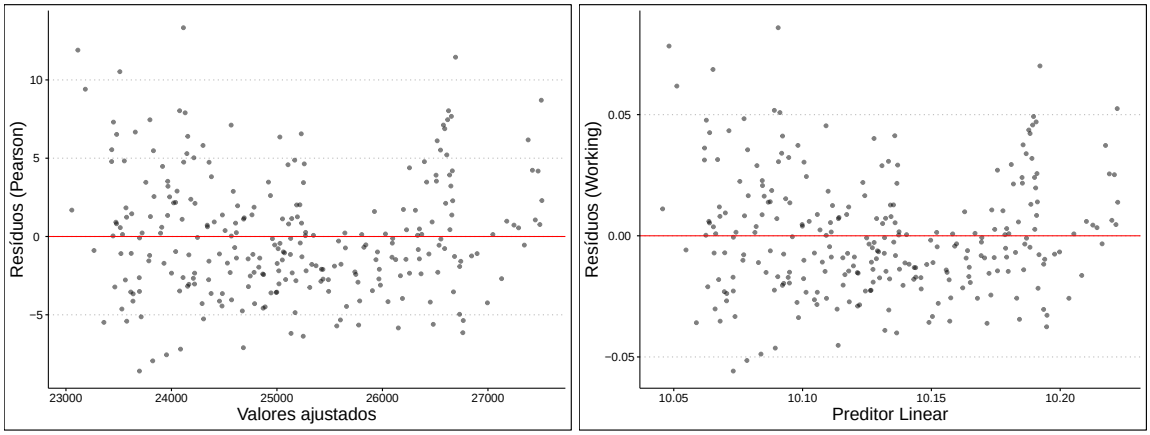


Figura 6 – Resíduos de Pearson *versus* Valores ajustados (esquerda) e Resíduos de trabalho (*working*) *versus* Valores ajustados (direita) para o Modelo Poisson Misto.

Para um breve diagnóstico do ajuste para o Modelo Poisson Misto, os gráficos das Figuras 5 e 6 foram construídos para diagnosticar uma possível falta de ajuste, avaliar a aleatoriedade dos resíduos e, principalmente, a sobredispersão, que é um problema comum nas aplicações de modelos de Poisson.

A Figura 5 indica que o modelo possui sobredispersão, posto que no primeiro há fuga de envelope tanto no gráfico seminormal, quanto para o QQ-Plot. Já no gráfico de Resíduos (Pearson) *versus* Valores ajustados apresentado na Figura 6, há resíduos com valores altos e, principalmente, no gráfico seminormal há uma exorbitante fuga do envelope. Por fim, o gráfico Resíduos de trabalho *versus* Valores ajustados, que são úteis para avaliar a adequação da função de ligação utilizada e, também, para identificação de *outliers*, exibe um comportamento de resíduos de trabalho, indicando que a escolha da função de ligação para o modelo parece adequada.

Tabela 6 – Estatísticas globais da qualidade de ajuste dos modelos Binomial Negativa.

Modelo	ω	AIC	BIC	logLik	Deviance	χ^2	$\neq$DF	Pr($>\chi^2$)
2	8	4267,2	4295,9	$-2125,6$	4251,2			
1	11	4267,4	4306,9	$-2122,7$	4245,4	5,7967	3	0,1219

4.1.3 Modelo Binomial Negativa

Na tentativa de corrigir o problema iminente da sobredispersão apresentada pelo modelo de Poisson, uma abordagem utilizando o modelo Binomial Negativa foi realizada. Embora o preditor linear utilizado seja o mesmo do modelo Poisson, a parametrização da variância condicional utilizada é

$$\mathbb{V}(Y_{ij}|\mathbf{b}_i) = \mathbb{E}(Y_{ij}|\mathbf{b}_i) + \frac{(\mathbb{E}(Y_{ij}|\mathbf{b}_i))^2}{\theta}$$

tal que θ é o parâmetro de dispersão, relacionado a ϕ da expressão (2.14) por meio da relação $\theta = 1/\phi$. Essa modificação ocorre para condizer com os métodos computacionais realizados pelo R (R Core Team, 2022) para realização do ajuste da Binomial Negativa.

Seguindo o mesmo raciocínio dos modelos anteriores, foram realizados dois ajustes, considerando a matriz de variância-covariância $\mathbf{\Sigma}$ sendo uma matriz simétrica positiva-definida qualquer ou uma matriz diagonal. Os resultados das estatísticas globais dos dois ajustes estão descritos na Tabela 6.

É possível constatar que a simplificação efetuada na estrutura da matriz de variância-covariância pode ser adotada, posto que os critérios informativos possuem diferenças pequenas e o teste Qui-Quadrado resultou num valor-p maior que $0{,}10$. Então, o Modelo 2, com menos parâmetros, é mais parcimonioso. As informações sobre as estimativas dos parâmetros estão na Tabela 7, na qual chama atenção o valor da estimativa do parâmetro de dispersão θ, responsável pela modelagem de variância extra em relação ao modelo Poisson. Na sequência, esse modelo será referido por Modelo Binomial Negativa Misto.

Tabela 7 – Estimativas dos parâmetros do Modelo Binomial Negativa Misto, e respectivas significâncias para os efeitos fixos

	Efeitos Fixos			
Parâmetro	Estimativa	Erro Padrão	Valor z	Pr($>\|z\|$)
β_0	$-9{,}039$	$0{,}006$	$-1466{,}089$	≈ 0
β_1	$0{,}0035$	$0{,}007$	$0{,}550$	$0{,}583$
β_2	$-0{,}053$	$0{,}003$	$-17{,}232$	≈ 0
β_3	$0{,}039$	$0{,}011$	$3{,}653$	$0{,}000259$
	Parâmetros de Variância			
	σ_0	σ_1	σ_2	θ
Estimativa	$0{,}006013$	$0{,}014281$	$0{,}005164$	$1714{,}868$

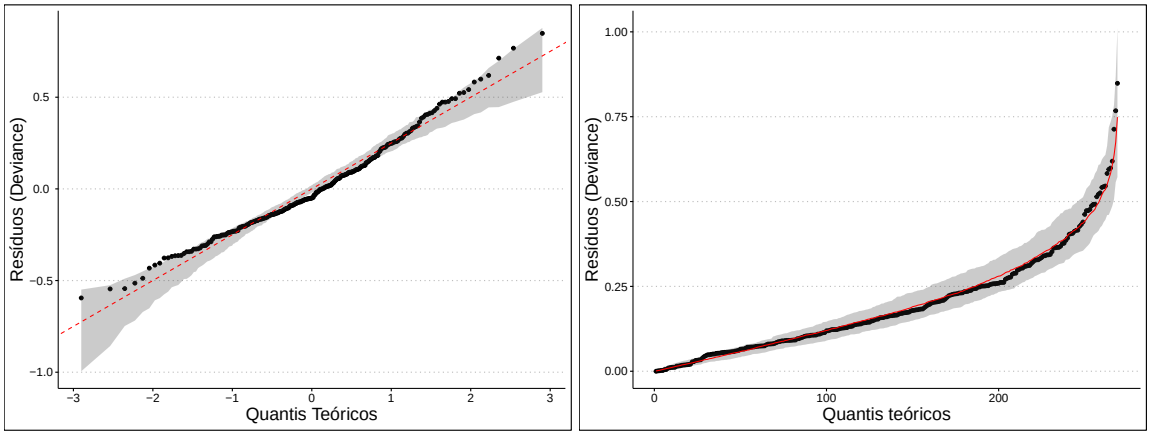


Figura 7 – QQ-Plot dos resíduos tipo Deviance (esquerda) e Gráfico Seminormal dos resíduos tipo Deviance (direita) para o Modelo Binomial Negativa Misto.

As estimativas dos efeitos fixos do Modelo Binomial Negativa Misto acompanham, em sinal, o mesmo comportamento já descrito para o Modelo Normal Misto. As estimativas dos componentes de variância são bem pequenas, indicando que o próprio modelo Binomial Negativa, com seu parâmetro de dispersão, já captura boa parte da heterogeneidade entre os anos.

Efetuando o diagnóstico para o Modelo Binomial Negativa Misto, os gráficos QQ-Plot e Seminormal apresentados na Figura 7 mostram um bom comportamento dos resíduos com relação à fuga do envelope, indicando uma correção efetiva da sobredispersão exorbitante apresentada pelo Modelo Poisson Misto. Já na Figura 8, tanto o gráfico de Resíduos de trabalho *Preditor Linear*, quanto o gráfico de Resíduos (Pearson) *versus* Valores ajustados, exibem um padrão aceitável, indicando que a função de ligação escolhida está adequada e que não há falta de ajuste.

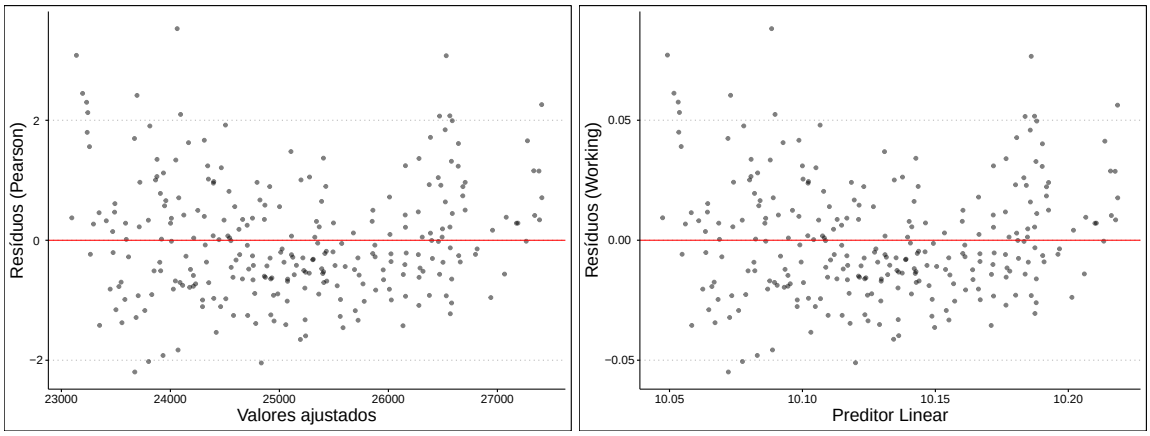


Figura 8 – Resíduos *versus* Valores ajustados (esquerda) e Resíduos (*Working*) *versus* Valores ajustados (direita) para o Modelo Binomial Negativa Misto.

4.2 Estimação da Mortalidade Esperada e Excesso de Mortes no Período Pandêmico

Utilizando o Modelo Normal Misto, a estimação da mortalidade esperada no período da 10^a à 52^a semana epidemiológica do ano 2020 (período pandêmico) foi dada pela aplicação da equação

$$\hat{\mathbb{E}}(Y_{6j}|\tilde{\mathbf{b}}_6) = (24195{,}62 + 199{,}06) + (138{,}23 + 405{,}23)\sin\left(\frac{2\pi}{52}x_j\right) + \\ + (-1367{,}60 + 186{,}48)\cos\left(\frac{2\pi}{52}x_j\right) + 1923{,}53t_{6j}$$

nas semanas sem observação, tal que $t_{6j} \in \left\{\frac{270}{269}, \frac{271}{269}, \ldots, \frac{312}{269}\right\}$ e $x_j = j$ para $j = 10,\ 11,\ \ldots,\ 52$. Já para o ano de 2021, a estimação da mortalidade foi dada pela aplicação da expressão

$$\hat{\mathbb{E}}(Y_{7j}|\tilde{\mathbf{b}}_7) = (24195{,}62 + \tilde{b}_{70}) + (138{,}23 + \tilde{b}_{71})\sin\left(\frac{2\pi}{52}x_j\right) + \\ + (-1367{,}60 + \tilde{b}_{72})\cos\left(\frac{2\pi}{52}x_j\right) + 1923{,}53t_{7j}$$

em que $\tilde{b}_{70} \sim \mathcal{N}(0; 199{,}06^2)$, $\tilde{b}_{71} \sim \mathcal{N}(0, 405.23^2)$, $\tilde{b}_{72} \sim \mathcal{N}(0, 186.48^2)$, $t_{7j} \in \left\{\frac{313}{269}, \frac{314}{269}, \cdots, \frac{364}{269}\right\}$ e $x_j = j$ para $j = 1, 2, \cdots, 52$.

Para a construção dos intervalos de predição a 95% para o ano de 2020, foi utilizado o processo de simulação conforme descrito na Seção 2.2.5, para $N = 1000$ gerações de σ^2 e do vetor de coeficientes da equação, a partir da distribuição normal multivariada, com média dada pela equação, em cada semana, e respectiva matriz de variância e covariância. Para contemplar a incerteza cometida em novas observações, para cada geração acima, amostrou-se Y a partir da normal univariada com variância σ^2. Por fim, para os intervalos de semana a semana, utilizou-se os percentis de ordem $2{,}5\%$ e de ordem $97{,}5\%$ das predições simuladas, além da necessidade, para o ano de 2021, simular os efeitos aleatórios para um novo grupo.

Por outro lado, utilizando o Modelo Binomial Negativa Misto, a estimação esperada no período da 10^a à 52^a semana do ano de 2020 foi dada pela utilização da equação

$$\hat{\mathbb{E}}(Y_{6j}|\tilde{\mathbf{b}}_6) = \exp\Big[(-9{,}039206 + 0{,}006013) + (0{,}003567 + 0{,}014281)\sin\left(\frac{2\pi}{52}x_j\right) + \\ +(-0{,}053919 + 0{,}005164)\cos\left(\frac{2\pi}{52}x_j\right) + 0{,}038614t_{6j}\Big] \times \text{pop}_6$$

nas semanas sem observação, $t_{6i} \in \left\{\frac{270}{269}, \frac{271}{269}, \cdots, \frac{312}{269}\right\}$, $x_j = j$ para $j = 10, 11, \cdots, 52$ e pop_6 é a projeção da população brasileira no ano 2020. Enquanto que para 2021, foi aplicada a expressão

$$\hat{\mathbb{E}}(Y_{7j}|\tilde{\mathbf{b}}_7) = \exp\Big[(-9{,}039206 + \tilde{b}_{70}) + (0{,}003567 + \tilde{b}_{71})\sin\left(\frac{2\pi}{52}x_j\right) + \\ +(-0{,}053919 + \tilde{b}_{72})\cos\left(\frac{2\pi}{52}x_j\right) + 0{,}038614t_{7j}\Big] \times \text{pop}_7,$$

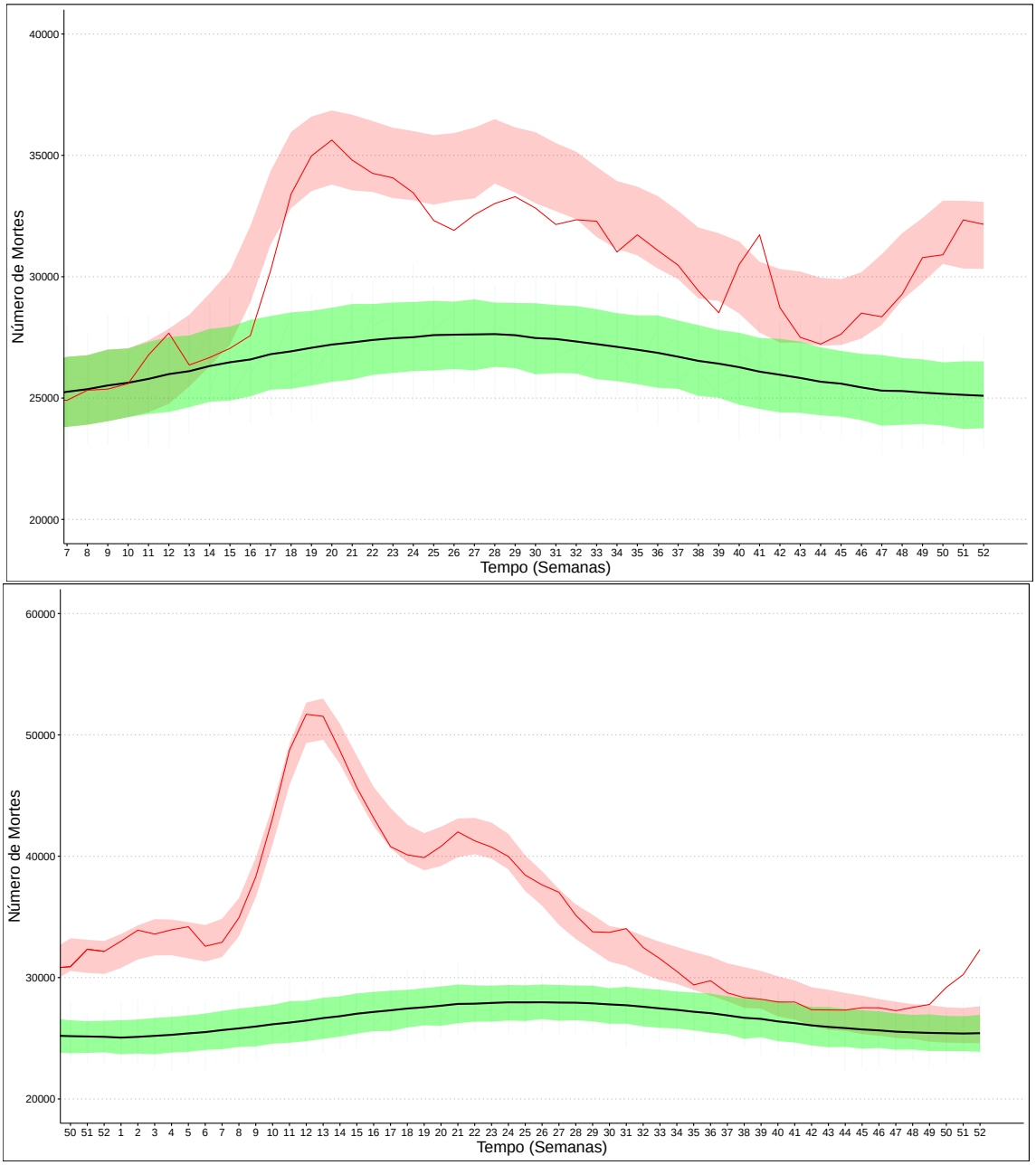


Figura 9 – Estimativas e intervalos de predição a 95% (Modelo Normal Misto), para a mortalidade esperada (faixa verde), mortalidade esperada + mortalidade por COVID-19 no período pandêmico (faixa vermelha) e mortalidade observada (linha vermelha) para o ano de 2020 (superior) e de 2021 (inferior).

em que $\tilde{b}_{70} \sim \mathcal{N}(0;\ 0{,}006013^2)$, $\tilde{b}_{71} \sim \mathcal{N}(0;\ 0{,}014281^2)$, $\tilde{b}_{72} \sim \mathcal{N}(0;\ 0{,}005164^2)$, $t_{7j} \in \left\{\frac{313}{269}, \frac{314}{269}, \cdots, \frac{364}{269}\right\}$, $x_j = j$ para $j = 1, 2, \cdots, 52$ e pop_7 é a projeção da população no ano de 2021.

Para a construção dos intervalos de predição a 95%, também foi utilizado o processo simulatório para $N = 1000$ gerações do vetor de coeficientes da equação, a partir da distribuição Normal multivariada, resultando em 1000 resultados da equação acima. A seguir, a partir da distribuição Binomial Negativa, gerou-se a predição de Y, com média dada pela equação, em cada

semana, e o parâmetro de dispersão estimado associado. Por fim, para representar os intervalos de semana a semana, foi empregado os percentis de ordem $2{,}5\%$ e $97{,}5\%$ das predições simuladas. Ademais, para o ano de 2021, foi necessária a estimação dos efeitos aleatórios para um novo grupo, conforme a Seção 2.2.5.

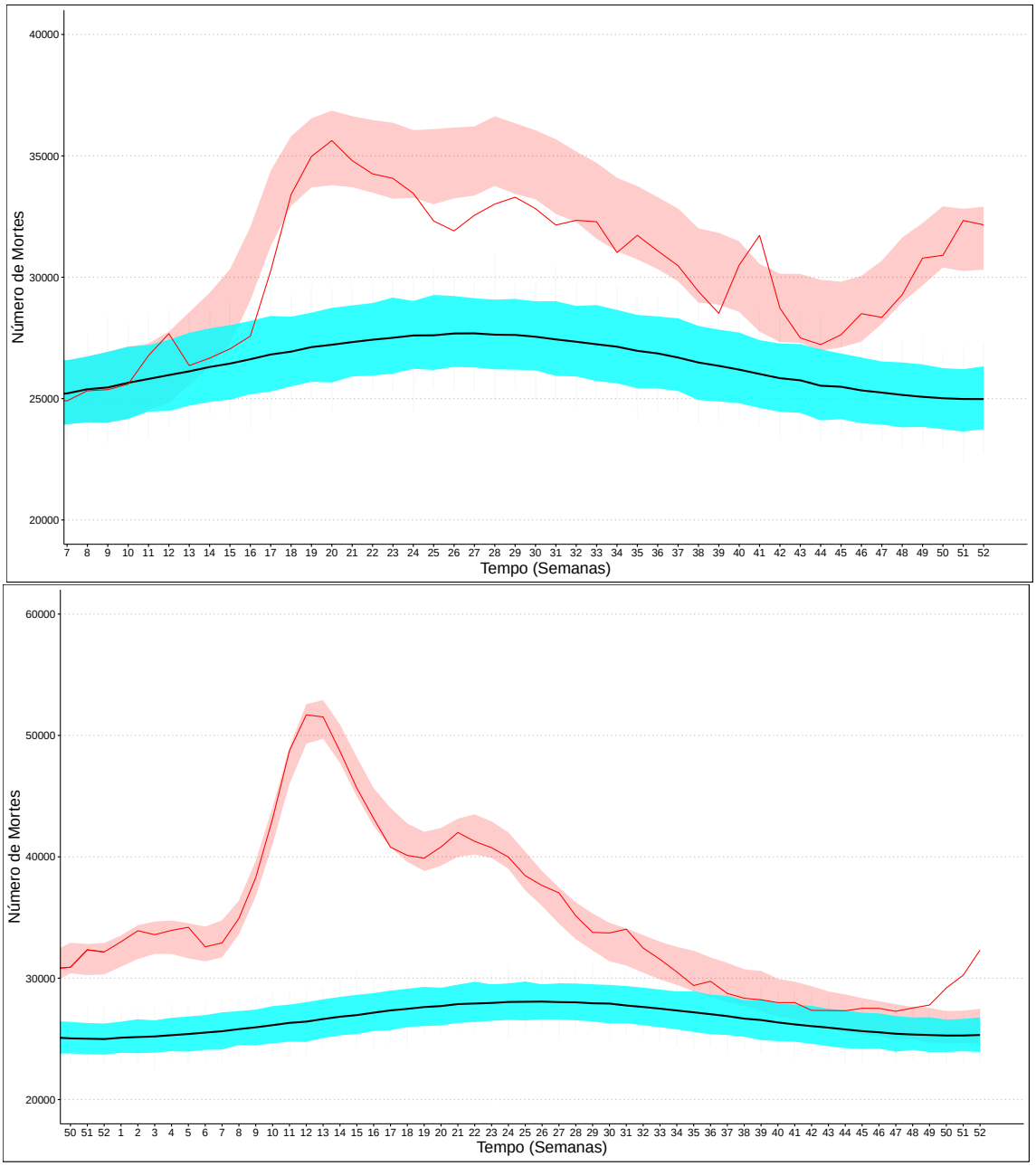


Figura 10 – Estimativas e intervalos de predição a 95% (Modelo Binomial Negativa Misto), para a mortalidade esperada (faixa azul), mortalidade esperada + mortalidade por COVID-19 no período pandêmico (faixa vermelha) e mortalidade observada (linha vermelha) para o ano de 2020 (superior) e de 2021 (inferior).

Os resultados, para cada modelo, são apresentados nas Figuras 9 e 10. A cada valor predito, foi acrescentado o número de mortes confirmadas por COVID-19, resultando, nas figuras, as faixas verde e azul, respectivamente. Essas faixas são as predições da mortalidade, caso o

excesso de mortes no período fosse apenas consequência direta pela doença. Nota-se que ambos os modelos fornecem predições bem similares, inclusive com relação à precisão.

Em cada figura, a linha vermelha corresponde à mortalidade de fato observada, somando todas as causas de morte.

A Figura 11 representa o resultado das predições realizadas exibindo o excesso de mortes para cada modelo misto (Normal e Binomial Negativa) utilizado para o período pandêmico. Ao analisar o ano de 2020, é possível observar intervalos dos quais a mortalidade observada de COVID-19 esteve abaixo das linhas tracejadas de excesso de morte, indicando que o excesso de mortalidade se deveu não apenas às mortes causadas diretamente pelo coronavírus, mas também a outros fatores. Esses espaços correspondem ao início da pandemia (9ª à 12ª semana) e, no segundo semestre, entre o final de setembro (39ª semana) e começo de outubro (41ª semana). Alguns indícios que podem ser a explicação para esses picos são: subnotificação de mortes por COVID-19 no começo da pandemia (Universidade Ferderal de Minas Gerais, 2022) e a proteção ao risco de infecções por outras causas durante o período de restrição social aplicados (CNN, 2021).

Nota-se que, em boa parte do período analisado, houve um deficit no excesso de mortalidade estimado em relação às mortes por COVID-19. Isso é evidenciado pelo fato de a linha vermelha permanecer acima das linhas pontilhadas coloridas, indicando que, nesse intervalo, o coronavírus foi a principal causa de mortalidade. Uma possível explicação para esse fenômeno é o relaxamento das políticas de distanciamento social (MARTINS; GUIMARÃES, 2022).

No ano de 2021, além do aumento geral do excesso de mortalidade em comparação ao ano anterior, observa-se que as mortes por COVID-19 estão alinhadas com o excesso de mortes estimado pelos modelos. Isso sugere que o excesso de mortalidade observado pode ser atribuído, essencialmente, às mortes diretamente causadas pelo coronavírus. É importante destacar que, nos períodos em que o excesso de mortes foi ocasionado por fatores secundários, casos como o aumento da mortalidade materna devido à falta de cuidados pré-natal tiveram influência significativa (UNFPA, 2022).

No final do período analisado, entre as semanas 49 e 52, nota-se uma divergência evidente. As linhas do modelo estimam um excesso de mortes superior ao número de óbitos registrados por COVID-19. Esse excedente pode ser interpretado como resultado de um aumento nas mortes por fatores secundários, tais como atrasos ou dificuldades no acesso a tratamentos médicos, flexibilização para as festas de final de ano, subnotificação de casos de COVID-19 devido à escassez de testes, piora das condições de saúde da população vulnerável devido à crise econômica e aumento de acidentes, violências ou mortes evitáveis.

Ainda, a Figura 11 destaca, para o ano de 2021, um declínio no número de mortes por coronavírus ao longo do ano. Esse comportamento é possivelmente explicado pelas políticas de vacinação, iniciadas em meados de janeiro, com foco nas populações mais idosas e com

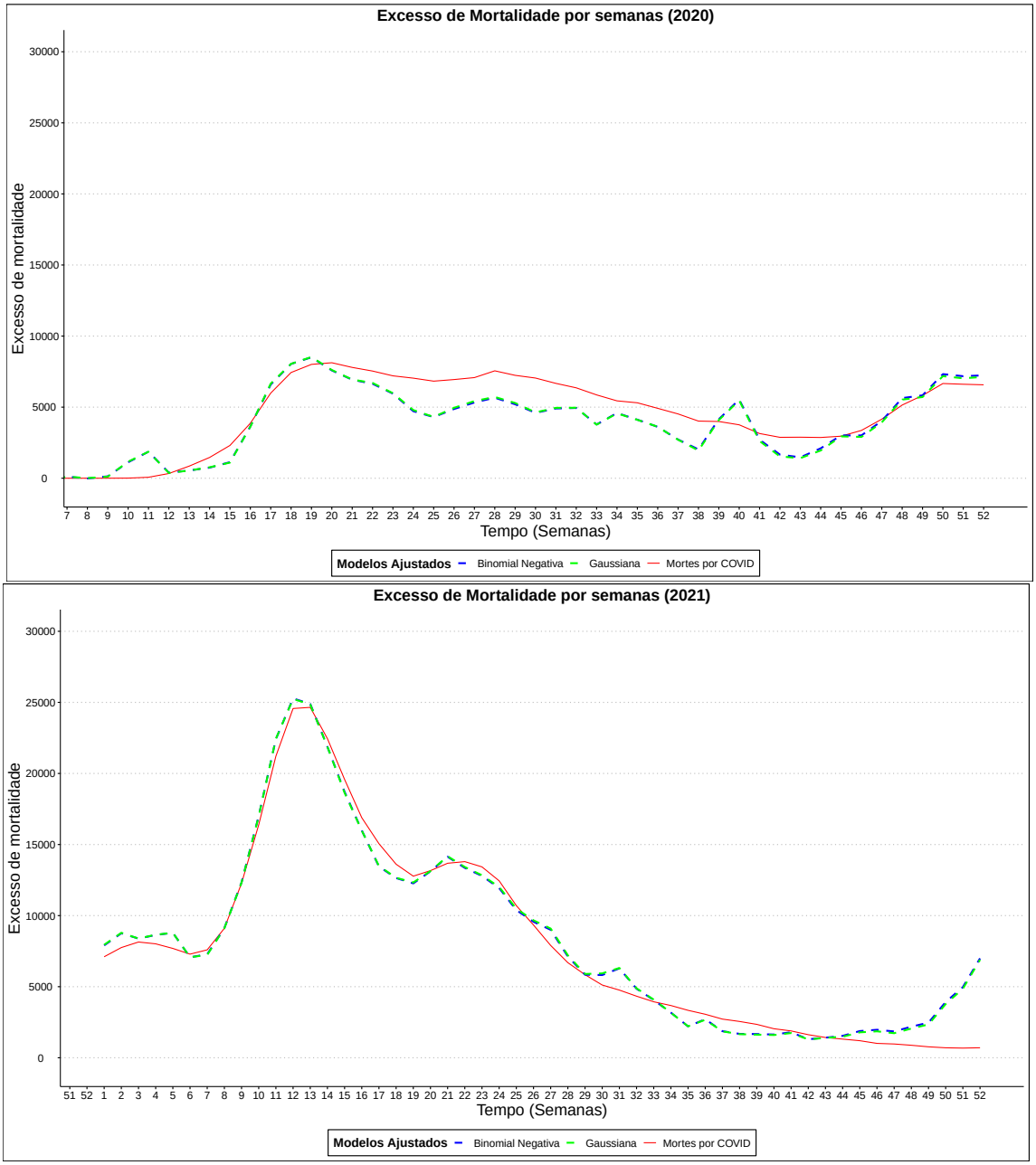


Figura 11 – Excesso de mortes utilizando os modelos mistos Normal (faixa pontilhada verde) e Binomial Negativa (faixa pontilhada em azul) e as mortes observadas por COVID-19 durante o ano de 2020.

comorbidades (Fiocruz, 2022).

5 Conclusão

Este trabalho propôs a aplicação de diferentes versões dos modelos generalizados mistos (Gaussiano, Poisson e Binomial Negativa) para estimar o excesso de mortalidade por COVID-19 no Brasil, destacando os desafios e vantagens associados a cada abordagem.

O estudo do excesso de mortalidade é fundamental para avaliar o impacto de eventos significativos, como a pandemia de COVID-19, identificando causas de morte diretas e indiretas. Assim, o desafio durante o trabalho foi estimar as mortes durante o período de pandemia sob a hipótese de não ocorrência, principalmente para os modelos de Poisson e Binomial Negativa, devido à falta de material em abundância sobre predições para novas observações na literatura e, também, para diagnósticos de modelos, quando comparada com os recursos disponíveis para o modelo Gaussiano. Para lidar com tais problemas, foi proposta a técnica de simulações, discutidas em Gelman e Hill (2007), para estimar a incerteza cometida ao realizar as predições, que demonstrou ser uma solução prática para lidar com MLGMs, posto que é uma escapatória para métodos numéricos mais sofisticados.

Durante os ajustes, dentre os três modelos propostos, o de Poisson foi o mais insatisfatório devido à sua insensibilidade com relação à sobredispersão, que demonstrou ser acentuada. Entretanto, o modelo Binomial Negativa, mesmo que com $\theta = 1714{,}87$, valor este que chama atenção durante o ajuste, conseguiu lidar com a variabilidade extra apresentada pelos dados, demonstrando ser um bom modelo alternativo ao de Poisson para dados de contagem. Quando o modelo Binomial Negativa é comparado com o Gaussiano, é possível notar que as estimativas de ambos são próximas. Entretanto, tanto na bibliografia teórica quanto na computacional, é possível realizar ajustes que melhoram o desempenho do modelo Normal, como flexibilização nas estruturas de variância e covariância. Como este não foi o foco do trabalho, este tópico não foi abordado.

Após realizados os diagnósticos dos modelos ajustados, foram selecionados o Modelo Gaussiano Misto e o Modelo Binomial Negativa Misto como sendo os parcimoniosos para prosseguirem com a análise da mortalidade.

A análise da mortalidade brasileira no período da 9ª semana de 2020 até o final de 2021 mostrou que existiu um grande excesso de mortes, principalmente em março de 2021, que foi estimado próximo de 25.000 mortes o excesso de mortalidade e que grande parte desse excesso

pode ser essencialmente atribuída ao coronavírus. Em 2020, picos do excesso de mortes podem ser explicados pelas políticas de flexibilização e subnotificações.

Esta pesquisa limitou-se à aplicação de modelos lineares generalizados (Normal, Poisson e Binomial Negativa), utilizando técnicas amplamente discutidas na literatura, além da adaptação de rotina computacional para simulações sob o Modelo Binomial Negativa Misto. Além disso, o estudo focou na mortalidade no Brasil como um todo, sem explorar fatores sociodemográficos disponíveis no conjunto de dados fornecido pelo SIM (Ministério da Saúde, 2023c).

Por fim, este trabalho contribui para o avanço das técnicas de predição no estudo do excesso de mortalidade, ao empregar métodos de complexidade intermediária que se situam entre modelos de regressão linear simples e técnicas avançadas de aprendizado de máquina.

Referências

BAEY, C.; COURNÈDE, P.-H.; KUHN, E. Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, v. 135, p. 107–122, 2019. 21

BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015. 9, 31

CNN. Fiocruz: Falta de leitos para covid é a ponta do iceberg do colapso da saúde. Notícias. 2021. Disponível em: <https://www.cnnbrasil.com.br/saude/fiocruz-falta-de-leitos-para-covid-e-a-ponta-do-iceberg-do-colapso-da-saude/>. 44

COLONIA, S. R. R. et al. Assessing covid-19 pandemic excess deaths in brazil: Years 2020 and 2021. *PLOS ONE*, v. 18, p. 1–26, 2023. 1, 30

DUNN, P.; SMYTH, G. *Generalized Linear Models With Examples in R*. New York: Springer, 2018. (Springer Texts in Statistics). 27

FAHRMEIR, L. et al. *Regression: Models, Methods and Applications*. Berlin Heidelberg: Springer-Verlag, 2013. 7, 17, 18

FARAWAY, J. *Linear Models with R*. Boca Raton: CRC Press, 2014. (Chapman & Hall/CRC Texts in Statistical Science). 5

FARAWAY, J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. New York: CRC Press, Taylor & Francis Group, 2016. (A Chapman & Hall Book). 7

Fiocruz. Brasil celebra um ano da vacina contra a covid-19. Notícias. 2022. Disponível em: <https://portal.fiocruz.br/noticia/brasil-celebra-um-ano-da-vacina-contra-covid-19?> 45

FITZMAURICE, G.; LAIRD, N.; WARE, J. *Applied Longitudinal Analysis*. Haboken: Wiley, 2012. (Wiley Series in Probability and Statistics). 7, 14, 16, 17, 18, 20, 25

FOX, J.; WEISBERG, S. *An R Companion to Applied Regression*. Los Angeles: SAGE Publications, 2011. 4

GELMAN, A.; HILL, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, 2007. (Analytical Methods for Social Research). 27, 28, 46

GELMAN, A.; SU, Y.-S. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. Vienna, Austria, 2024. R package version 1.14-4. Disponível em: <https://CRAN.R-project.org/package=arm>. 31

GUMEDZE, F.; DUNNE, T. Parameter estimation and inference in the linear mixed model. *Linear Algebra and its Applications*, v. 435, n. 8, p. 1920–1944, 2011. 20

HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, v. 54, n. 1-2, p. 93–108, 1967. 8

HENDERSON, C. R. Estimation of variance and covariance components. *Biometrics*, v. 9, n. 2, p. 226–252, 1953. 8

HILDEN-MINTON, J. A. *Multilevel diagnostics for mixed and hierarchical linear models*. Tese (Doutorado) — University of California, LA, 1995. 26

HUI, F. K. C.; MÜLLER, S.; WELSH, A. H. Random effects misspecification can have severe consequences for random effects inference in linear mixed models. *International Statistical Review*, v. 89, n. 1, p. 186–206, 2021. 20

Instituto Brasileiro de Geografia e Estatística. *População Brasileira*. 2024. Dados sobre a população brasileira anualmente. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html?=&t=resultados>. 32

JIANG, J.; NGUYEN, T. *Linear and Generalized Linear Mixed Models and Their Applications*. New York: Springer, 2021. (Springer Series in Statistics). 19, 20

KENWARD, M. G.; ROGER, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, v. 53, n. 3, p. 983–997, 1997. 22

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. *Biometrics*, v. 38, n. 4, p. 963–974, 1982. 8

LUKE, S. G. Evaluating significance in linear mixed-effect models in r. *Behavior research methods*, v. 49, p. 1494–1502, 2017. 21

MALTA, M. et al. Political neglect of covid-19 and the public health consequences in brazil: The high costs of science denial. *EClinicalMedicine*, v. 35, p. 100878, 2021. 1

MARTINS, T. C. d. F.; GUIMARÃES, R. M. Distanciamento social durante a pandemia da covid-19 e a crise do estado federativo: um ensaio do contexto brasileiro. *Saúde em Debate*, v. 46, p. 265—280, 2022. 44

MCCULLOCH, C.; SEARLE, S.; NEUHAUS, J. *Generalized, Linear, and Mixed Models*. New York: Wiley, 2011. (Wiley Series in Probability and Statistics). 7, 18, 20

Ministério da Saúde. *Apresentação SIM*. 2023. Acessado em 13/07/2024. Disponível em: <https://svs.aids.gov.br/daent/cgiae/sim/apresentacao/>. 32

Ministério da Saúde. *Painel Coronavírus*. 2023. Disponível em: <https://covid.saude.gov.br/>. 1

Ministério da Saúde. *Sistema de Informação sobre Mortalidade*. 2023. Disponível em: <https://opendatasus.saude.gov.br/dataset/sim>. 32, 47

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-normal plots and overdispersed models in R: The hnp package. *Journal of Statistical Software*, v. 81, n. 10, p. 1–23, 2017. 31

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, v. 135, n. 3, p. 370–384, 1972. 5

PALACIO-MEJÍA, L. S. et al. Leading causes of excess mortality in mexico during the covid-19 pandemic 2020-2021: A death certificates study in a middle-income country. *The Lancet Regional Health – Americas*, p. 100303, 2022. 30

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, v. 58, n. 3, p. 545–554, 1971. 8, 12

PAULA, G. A. *Modelos de Regressão: com apoio computacional*. São Paulo: IME-USP, 2013. 5, 7

PINHEIRO, J.; BATES, D. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag, 2009. (Statistics and Computing). 7, 9, 21, 22

PINHEIRO, J.; BATES, D.; R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. Vienna, Austria, 2023. R package version 3.1-163. Disponível em: <https://CRAN.R-project.org/package=nlme>. 31

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <https://www.R-project.org/>. 31, 39

SANTOS, A. M. d. et al. Excess deaths from all causes and by covid-19 in brazil in 2020. *Revista de Saúde Pública*, v. 55, p. 71, 2021. 30, 32

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, v. 2, n. 6, p. 110–114, 1946. 22

SEARLE, S.; CASELLA, G.; MCCULLOCH, C. *Variance Components*. Haboken: Wiley, 2006. 13

SINGER, J. M.; ROCHA, F. M.; NOBRE, J. S. Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review*, v. 85, n. 2, p. 290–324, 2017. 24, 25, 26, 31, 34

STRAM, D. O.; LEE, J. W. Variance components testing in the longitudinal mixed effects model. *Biometrics*, v. 50, n. 4, p. 1171–1177, 1994. 21

STRINGER, A.; BILODEAU, B. *Fitting Generalized Linear Mixed Models using Adaptive Quadrature*. 2022. 20

TUERLINCKX, F. et al. Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, v. 59, n. 2, p. 225–255, 2006. 14, 20, 22, 23, 24, 28

UNFPA. Unfpa: mortalidade materna no brasil aumentou 94,4% durante a pandemia. Notícias. 2022. Disponível em: <https://brasil.un.org/pt-br/203964-unfpa-mortalidade-materna-no-brasil-aumentou-944-durante-pandemia>. 44

Universidade Ferderal de Minas Gerais. Pesquisa estima pelo menos 18% de subnotificação de mortes por covid-19 no país. Notícias. 2022. Disponível em: <https://ufmg.br/comunicacao/noticias/pesquisa-estima-pelo-menos-18-de-subnotificacao-de-mortes-por-covid-19-no-pais?> 44

VERBEECK, J. et al. A linear mixed model to estimate COVID-19-induced excess mortality. *Biometrics*, 2021. 30

VERBEKE, G.; LESAFFRE, E.; BRANT, L. J. The detection of residual serial correlation in linear mixed models. *Statistics in Medicine*, v. 17, n. 12, p. 1391–1402, 1998. 25

WHO. *Methods for estimating the excess mortality associated with the COVID-19 pandemic*. Geneva, Switzerland, 2023. 1, 30

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016. 31

YATES, F. The recovery of inter-block information in balanced incomplete block designs. *Annals of Eugenics*, v. 10, n. 1, p. 317–325, 1940. 8