

USO DE POLINÔMIOS FRACIONÁRIOS NOS MODELOS MISTOS

EDIJANE PAREDES GARCIA

Tese apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para obtenção do título de Doutor em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2019

USO DE POLINÔMIOS FRACIONÁRIOS NOS MODELOS MISTOS

EDIJANE PAREDES GARCIA

Orientadora: Profa. Dra. **Luzia Aparecida Trinca**

Tese apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para obtenção do título de Doutor em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2019

Ficha Catalográfica

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Garcia, Edijane Paredes.

Uso de polinômios fracionários nos modelos mistos /
Edijane Paredes Garcia. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista
"Júlio de Mesquita Filho", Instituto de Biociências de
Botucatu

Orientador: Luzia Aparecida Trinca

Capes: 90194000

1. Curvatura. 2. Estudos longitudinais. 3. Polinômios.
4. Análise de regressão.

Palavras-chave: Curvatura; Dados longitudinais; Forma
funcional; Modelo polinomial fracionário; Seleção de
variáveis.

Dedicatória

À minha mãe Tereza, pelo apoio incondicional

À minha filha Liz Yamile, por existir na minha vida

Agradecimentos

Antes de tudo agradeço a Deus pela presença constante em minha vida, por me guiar com muita luz e determinação para superar as dificuldades, pelas minhas conquistas e crescimento com a realização deste trabalho.

Agradeço em especial à minha mãe, Maria Tereza, e aos meus irmãos, Eneilane e Junio, por todo o incentivo, amor e incondicional apoio para que pudesse vir para o doutorado.

Um agradecimento muito especial à minha filha, Liz Yamile, razão da minha vida e minha companheira há 9 anos, por ser tão responsável, independente e ter contribuído com aquele abraço e aquele EU TE AMO, inesperado, tão gostoso e tão importante para eu continuar em frente.

Meus mais sinceros agradecimentos à minha orientadora, Luzia Aparecida Trinca, que desde o momento que aceitou me orientar, motivou-me mais ainda a crescer na área de meu interesse. Agradeço pela orientação, apoio, paciência, compreensão, confiança em mim depositada, além de ter proporcionado as melhores condições para desenvolver o meu trabalho. Sinto-me honrada em poder dizer que trabalhamos juntas.

Aos professores componentes da minha banca examinadora, Luzia Aparecida Trinca, Julio da Motta Singer, Juvêncio Santos Nobre, Silvio Sandoval Zocchi e Miriam Harumi Tsunemi pelas valiosas contribuições que colaboraram para o enriquecimento deste trabalho.

À minha professora Rosana Cristina Pereira Parente, minha segunda mãe, uma pessoa mais que incrível, indescritível, capaz de tudo ao seu alcance para ver o meu sucesso. Agradeço-a pela maravilhosa profissional e pessoa que é e sempre foi, à sua

infinita paciência, ao seu carisma, à sua sabedoria.

Aos professores do Departamento de Estatística da Universidade Federal do Amazonas, em especial aos professores Celso Rômulo Barbosa Cabral, por acreditar que sou capaz, José Cardoso Neto, meu grande incentivador, Max Sousa de Lima, pelo brilhante parecer para minha saída para o doutorado e Themis da Costa Abensur, por se fazer sempre presente e estar na torcida pela minha vitória.

Aos professores do Departamento de Bioestatística da UNESP/Botucatu que contribuíram para a minha formação e, em especial, ao Prof. Dr. Carlos Alberto Padovani com quem aprendi muito e de quem sentirei muita falta pelas boas risadas e conversas na sala 8.

Aos funcionários do Departamento de Bioestatística da UNESP/Botucatu Junior, Luiz, Arthur e Dona Selda, uma equipe de verdade, todos sempre prontos a ajudar.

À minha *best friend* Odete dos Santos Amaral e toda a sua linda família, Dona Odete, Piedade, Claudete, Telma, Marise e Marcelle, por me presentarem com seus inestimáveis apoios em distintos momentos e por suas presenças afetivas, o meu reconhecido e carinhoso muito obrigada.

Aos colegas da turma de doutorado pela convivência e conhecimentos compartilhados e, em especial, às minhas amigas Livia Paschoalino de Campos, Rita de Cássia de Lima Idalino e Magali Teresópolis Reis Amaral pela força nos períodos difíceis e pelos muitos momentos de descontração.

Ao meu eterno amigo, Perseverando da Trindade Garcia Filho, por todo suporte no início desta jornada.

Às *Lokas Poderosas* Solange Huber dos Santos, Núbia Rodrigues de Souza, Izaura Jardim, Jeane Cristina Ribeiro Lima e Yara Cláudia Renovato dos Santos, pelo apoio e incentivo durante todos os anos no doutorado.

Sumário

	Página
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xiv
RESUMO	xvii
SUMMARY	xix
1 INTRODUÇÃO	1
1.1 Objetivos	4
1.2 Organização da tese	5
1.3 Suporte computacional	5
2 POLINÔMIOS FRACIONÁRIOS	7
2.1 Introdução	7
2.2 Polinômios fracionários para uma variável regressora contínua	8
2.2.1 Mudança da origem da regressora	15
2.2.2 Procedimento de seleção da função	16
2.3 Polinômios fracionários para várias variáveis regressoras	19
2.3.1 Algoritmo MFP	19
2.4 Extensão do procedimento MFP para investigar interações	21
2.4.1 Algoritmo MFPI	22
3 MODELOS LINEARES MISTOS	24
3.1 Especificação do modelo	25
3.2 Estimação dos parâmetros	27
3.3 Estrutura para a matriz de covariâncias \mathbf{R}_i	29

3.3.1	Estrutura de variância	30
3.3.2	Estrutura de correlação	33
3.3.3	Procedimento para seleção da estrutura de covariância \mathbf{R}_i	35
3.4	Testes para inclusão de efeitos aleatórios	38
3.4.1	Teste de permutação baseado nos BLUPs	40
3.4.2	Teste de permutação baseado na RVRG	43
3.5	Diagnóstico	45
3.5.1	Resíduos marginais	47
3.5.2	Resíduos condicionais	47
3.5.3	Resíduos de efeitos aleatórios	49
3.5.4	Análise de pontos alavanca	50
3.5.5	Análise de omissão de casos	53
3.6	Outras distribuições para os efeitos aleatórios	55
4	POLINÔMIOS FRACIONÁRIOS NOS MODELOS MISTOS	57
4.1	Introdução	57
4.2	O Modelo PFMM	59
4.3	Estratégia de ajuste do modelo PFMM	61
4.3.1	Etapa 1: Seleção de uma estrutura média preliminar	61
4.3.2	Etapa 2: Seleção dos efeitos aleatórios	63
4.3.3	Etapa 3: Seleção de uma estrutura de covariância para os erros	64
4.3.4	Etapa 4: Redução do modelo	65
5	APLICAÇÕES	66
5.1	Introdução	66
5.2	Aplicação 1: Concentração da bilirrubina em recém-nascidos nos primeiros doze dias de vida	66
5.2.1	Modelo PFMM	68
5.2.1.1	Etapa 1: Seleção de uma estrutura média preliminar	68
5.2.1.2	Etapa 2: Seleção dos efeitos aleatórios	71
5.2.1.3	Etapa 3: Seleção de uma estrutura de covariâncias para os erros	73

5.2.1.4	Etapa 4: Redução do modelo	85
5.3	Aplicação 2: Dados da taxa de ultrafiltração de dialisadores de alto fluxo	91
5.3.1	Descrição do estudo	91
5.3.2	Modelo linear misto	92
5.3.3	Modelo PFMM	94
5.3.3.1	Etapa 1: Seleção de uma estrutura média preliminar	95
5.3.3.2	Etapa 2: Seleção dos efeitos aleatórios	96
5.3.3.3	Etapa 3: Seleção de uma estrutura de covariância para os erros	98
5.3.3.4	Etapa 4: Redução do modelo	103
5.4	Aplicação 3: Efeitos da poluição sobre o preço de residências	109
5.4.1	Descrição do estudo	109
5.4.2	Modelo de regressão múltipla	113
5.4.3	Modelo linear misto	117
5.4.4	Modelo PFMM	118
5.4.4.1	Etapa 1: Seleção de uma estrutura média preliminar	118
5.4.4.2	Etapa 2: Seleção dos efeitos aleatórios	126
5.4.4.3	Etapa 3: Seleção de uma estrutura de covariância para os erros	126
5.4.4.4	Etapa 4: Redução do modelo	131
6	DISCUSSÕES E CONCLUSÕES	141
6.1	Pesquisas futuras	142
	ANEXOS	144
	REFERÊNCIAS	155

Lista de Figuras

	Página
2.1 Ilustração das três curvas FP1 com potências $p < 0$ e coeficiente $\beta_1 < 0$. . .	13
2.2 Ilustração das cinco curvas FP1 com potências $p \geq 0$ e coeficiente $\beta_1 > 0$. . .	14
2.3 Ilustração de seis curvas FP2 com diferentes potências (p_1, p_2) e coeficientes (β_1, β_2) . Números entre parênteses indicam as potências.	15
5.1 Perfil médio.	67
5.2 Perfis individuais.	68
5.3 Gráfico da componente do desvio em função da potência $p \in S$ para o ajuste do modelo em (5.1).	69
5.4 Gráfico da componente do desvio em função das potências $(p_1, p_2) \in S^2$ para o ajuste do modelo em (5.2).	69
5.5 Gráfico do perfil médio dos dados observados e curvas ajustadas pelos valores estimados do modelo PFMM2 (5.3) e modelo polinomial cúbico com efeitos mistos (PCMM).	71
5.6 Correlações em função da defasagem.	74
5.7 Variograma amostral.	75
5.8 Dados observados (pontos) e curvas de predição para os modelos (5.4) e (2c).	77
5.9 Dados observados (pontos) e curvas de predição para o modelo (5.7). . .	81
5.10 Gráfico do resíduo condicional padronizado <i>versus</i> valor ajustado para o modelo em (5.7).	82
5.11 Gráfico do resíduo marginal padronizado <i>versus</i> valor ajustado para o modelo em (5.7).	82

5.12	Gráfico do envelope simulado com 95% para resíduos condicionais minimamente confundidos para o modelo em (5.7).	83
5.13	Gráfico do índice de Lesaffre-Verbeke modificado <i>versus</i> índice do recém-nascido para o modelo (5.7).	83
5.14	Gráfico baseado na distância de Mahalanobis <i>versus</i> índice do recém-nascido para o modelo em (5.7).	84
5.15	Gráfico QQ baseado na distância de Mahalanobis correspondente ao modelo em (5.7).	84
5.16	Gráfico da componente do desvio para 36 modelos PFMM2(p_1, p_2) com efeitos aleatórios e estrutura de covariância do erro selecionados nas etapas 2 e 3, respectivamente.	85
5.17	Gráfico do perfil médio dos dados observados e curva ajustada pelos valores preditos do modelo em (5.8).	87
5.18	Gráfico da taxa de variação para a curva média marginal em (5.8) em função da idade.	88
5.19	Dados observados (pontos) e curvas de predição para o modelo (5.8). . .	89
5.20	Gráficos de diagnóstico para o modelo em (5.8).	90
5.21	Perfis individuais dos dialisadores segundo a taxa de ultrafiltração (ml/hr) medida em 7 níveis crescentes de pressão transmembrana (dmHg) para as taxas de fluxo de sangue de 200 dl/min e 300 dl/min.	92
5.22	Gráfico da componente do desvio estimada pelos modelos (5.11) e (5.12) em função da potência p_1 estratificada por p_2	95
5.23	Gráfico do resíduo condicional padronizado <i>versus</i> pressão transmembrana para o modelo em (5.15).	99
5.24	Gráfico do resíduo condicional padronizado em função da pressão transmembrana para o modelo em (5.16).	100
5.25	Gráfico do resíduo marginal padronizado <i>versus</i> valor ajustado para o modelo em (5.16).	101

5.26	Gráfico do envelope simulado com 95% para resíduos condicionais minimamente confundidos para o modelo em (5.16).	101
5.27	Gráfico do índice de Lesaffre-Verbeke modificado <i>versus</i> índice do dialisador para o modelo em (5.16).	102
5.28	Gráfico baseado na distância de Mahalanobis <i>versus</i> índice do dialisador para o modelo em (5.16).	102
5.29	Gráfico QQ baseado na distância de Mahalanobis correspondente ao modelo em (5.16).	103
5.30	Gráfico do perfil médio com curvas ajustadas pelos valores preditos do modelo em (5.20) por taxa de fluxo de sangue de 200 dl/min e de 300 dl/min.	106
5.31	Taxa de variação em função da pressão transmembrana para os níveis de fluxo de sangue de 200 dl/min e 300 dl/min.	106
5.32	Curvas individuais ajustadas da taxa de ultrafiltração em função da pressão transmembrana correspondentes aos modelos (5.9), PCMM, e (5.20), PFMM.	107
5.33	Gráficos de diagnóstico para o modelo em (5.20).	108
5.34	Mapa do estado de Massachusetts com seus 14 condados.	110
5.35	Gráficos dos resíduos padronizados <i>versus</i> as variáveis regressoras do modelo em (5.22).	115
5.36	Gráfico de probabilidades normal, com envelope gerado, para os resíduos padronizados do modelo em (5.22).	116
5.37	Boxplots dos resíduos brutos por localidade correspondente ao modelo em (5.22).	116
5.38	Gráficos para análise de resíduos referente ao modelo em (5.23).	118
5.39	Gráfico do índice de Lesaffre-Verbeke modificado <i>versus</i> os índices das localidades, modelo (5.24).	127
5.40	Gráfico dos resíduos condicionais padronizados <i>versus</i> os valores preditos, modelo (5.24).	127

5.41	Gráfico do resíduo minimamente confundido <i>versus</i> quantis da distribuição $N(0,1)$, modelo (5.24).	128
5.42	Gráfico do índice de Lesaffre-Verbeke modificado <i>versus</i> os índices das localidades, modelo (5.28).	130
5.43	Gráfico de probabilidades normal, com envelope gerado, para os resíduos minimamente confundidos padronizados, modelo (5.28).	130
5.44	Gráficos das distâncias de Mahalanobis, modelo (5.28).	131
5.45	Gráfico do resíduo marginal padronizado <i>versus</i> valor ajustado para o modelo em (5.29).	138
5.46	Gráfico do resíduo marginal padronizado <i>versus</i> unidade observacional para o modelo em 5.29.	138
5.47	Gráfico do envelope simulado com 95% para o resíduo com confundimento mínimo para o modelo em (5.29).	139
5.48	Gráfico do índice de Lesaffre-Verbeke modificado <i>versus</i> índice da localidade para o modelo em (5.29).	139
5.49	Gráficos baseado na distância de Mahalanobis <i>versus</i> índice da localidade para o modelo em (5.29).	140
5.50	Gráficos QQ baseado na distância de Mahalanobis correspondentes ao modelo em (5.29).	140
6.1	Gráfico da distância de Cook condicional para o modelo (5.8).	144
6.2	Gráfico da distância de Cook condicional $D1i$ para o modelo (5.8).	145
6.3	Gráfico da distância de Cook condicional $D2i$ para o modelo (5.8).	145
6.4	Gráfico da distância de Cook condicional $D3i$ para o modelo (5.8).	146
6.5	Gráfico da alavanca marginal generalizada L_i para o modelo (5.8).	146
6.6	Gráfico da alavanca marginal generalizada $L1i$ para o modelo (5.8).	147
6.7	Gráfico da alavanca marginal generalizada $L2i$ para o modelo (5.8).	147
6.8	Gráfico da distância de Cook condicional para o modelo (5.20).	148
6.9	Gráfico da distância de Cook condicional $D1i$ para o modelo (5.20).	149
6.10	Gráfico da distância de Cook condicional $D2i$ para o modelo (5.20).	149

6.11	Gráfico da distância de Cook condicional D_{3i} para o modelo (5.20). . . .	150
6.12	Gráfico da alavanca marginal generalizada L_i para o modelo (5.20). . . .	150
6.13	Gráfico da alavanca marginal generalizada L_{1i} para o modelo (5.20). . . .	151
6.14	Gráfico da alavanca marginal generalizada L_{2i} para o modelo (5.20). . . .	151
6.15	Gráfico da distância de Cook condicional para o modelo (5.29).	152
6.16	Gráfico da distância de Cook condicional D_{1i} para o modelo (5.29). . . .	153
6.17	Gráfico da distância de Cook condicional D_{2i} para o modelo (5.29). . . .	153
6.18	Gráfico da distância de Cook condicional D_{3i} para o modelo (5.29). . . .	154

Lista de Tabelas

	Página
2.1 Exemplos ilustrativos de funções FP2 para uma variável regressora . . .	10
2.2 Ponto de mínimo ou máximo em x em uma função FP2 unimodal dadas suas potências e seus coeficientes ($\kappa = -\beta_1/\beta_2$)	13
2.3 Comparando modelos FP1 e FP2	18
3.1 Exemplos de funções de variância que dependem de δ , mas não dependem de μ^a	32
3.2 Exemplos de funções de variância que dependem de δ e μ^a	32
3.3 Exemplos de funções de variância que dependem de μ , mas não de δ^a . .	32
3.4 Exemplos de estruturas de correlação serial	34
3.5 Tipos de gráficos de diagnóstico baseados nos resíduos marginais, con- dicionais e de efeitos aleatórios, propostos em Singer et al. (2017), para avaliar o ajuste do modelo linear misto	50
3.6 Tipos de gráficos de influência baseados na decomposição da distância de Cook condicional e na alavanca generalizada no contexto do modelo linear misto	55
5.1 Aplicação do procedimento FSP para selecionar o melhor modelo PFMM para dados da bilirrubina	70
5.2 Resultados dos testes de permutação e assintótico para inclusão de efeitos aleatórios específicos ao modelar a concentração de bilirrubina em RNs .	72
5.3 Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo em (5.4)	73

5.4	Estatísticas de ajuste do modelo (5.4) com modificações na estrutura de covariâncias do erro para três diferentes funções de variância e estruturas de correlação	76
5.5	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.7)	79
5.6	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.8)	86
5.7	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.9)	94
5.8	Estimativas de máxima verossimilhança e erros padrão estimados associados aos parâmetros do modelo em (5.14)	96
5.9	Resultados dos testes de permutação e assintótico para inclusão de efeitos aleatórios específicos ao modelar a taxa de ultrafiltração em dialisadores de alto fluxo	97
5.10	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo em (5.15)	98
5.11	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.20)	105
5.12	Observações da variável dependente que foram codificadas erroneamente	112
5.13	Estimativas de máxima verossimilhança, erro padrão e valor-p ordenados para os parâmetros de regressão do modelo (5.23)	119
5.14	Ciclo 1 do algoritmo MFP para seleção de uma estrutura média preliminar	123
5.15	Ciclo 2 do algoritmo MFP para seleção de uma estrutura média preliminar	124
5.16	Ciclo 3 do algoritmo MFP para seleção de uma estrutura média preliminar	125
5.17	Graus de liberdade, estatística F e valor-p ordenados para os parâmetros de regressão do modelo (5.28)	132
5.18	Ciclo 1 do algoritmo MFP para redução do modelo	134
5.19	Ciclo 2 do algoritmo MFP para redução do modelo	135

5.20	Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros dos modelos (5.23), (5.24), (5.28) e (5.29) . . .	136
------	---	-----

USO DE POLINÔMIOS FRACIONÁRIOS NOS MODELOS MISTOS

Autora: EDIJANE PAREDES GARCIA

Orientadora: Profa. Dra. LUZIA APARECIDA TRINCA

RESUMO

A classe dos modelos de regressão incorporando polinômios fracionários - FPs (*Fractional Polynomials*), proposta por Royston & Altman (1994), tem sido amplamente estudada. O uso de FPs em modelos mistos constitui uma alternativa muito atrativa para explicar a dependência das medidas intra-unidades amostrais em modelos em que há não linearidade na relação entre a variável resposta e variáveis regressoras contínuas. Tal característica ocorre devido aos FPs oferecerem, para a resposta média, uma variedade de formas funcionais não lineares para as variáveis regressoras contínuas, em que se destacam a família dos polinômios convencionais e algumas curvas assimétricas e com assíntotas. A incorporação dos FPs na estrutura dos modelos mistos tem sido investigada por diversos autores. Porém, não existem publicações sobre: a exploração da problemática da modelagem na parte fixa e na parte aleatória (principalmente na presença de várias variáveis regressoras contínuas e categóricas); o estudo da influência dos FPs na estrutura dos efeitos aleatórios; a

investigação de uma adequada estrutura para a matriz de covariâncias do erro; ou, um ponto de fundamental importância para colaborar com a seleção do modelo, a realização da análise de diagnóstico dos modelos ajustados. Uma contribuição, do nosso ponto de vista, de grande relevância é a investigação e oferecimento de estratégias de ajuste dos modelos polinômios fracionários com efeitos mistos englobando os pontos citados acima com o objetivo de preencher essas lacunas e de despertar aos usuários o imenso potencial dos modelos mistos, agora ainda mais ampliado, para a modelagem de dados correlacionados. Nesta tese propomos uma estratégia para a implementação do uso de FPs para modelar tendências não lineares em dados agrupados, tais como medidas repetidas, dados longitudinais e multiníveis. Dentro dessa proposta possibilitamos a inclusão de componentes de interação entre as transformações FP e variáveis categóricas por meio do uso de uma adaptação do algoritmo MFPI (*Multivariable Fractional Polynomial Interaction*), proposto por Royston & Sauerbrei (2004), realizamos a modelagem da matriz de covariâncias e utilizamos ferramentas gráficas para análise de diagnóstico (Singer et al., 2017) com o objetivo de verificar possíveis violações das suposições do modelo, bem como avaliamos o efeito de observações influentes nos resultados do ajuste. Concluímos com análises de três exemplos práticos em que ajustes sob os modelos mistos são comparados. No desenvolvimento dessa tese utilizamos a linguagem de programação R como suporte computacional.

Palavras-chave: Curvatura, dados longitudinais, forma funcional, modelo polinomial fracionário, seleção de variáveis.

USE OF FRACTIONAL POLYNOMIALS IN MIXED MODELS

Author: EDIJANE PAREDES GARCIA

Adviser: Profa. Dra. LUZIA APARECIDA TRINCA

SUMMARY

The class of regression models incorporating Fractional Polynomials (FPs), proposed by Royston & Altman (1994), has been extensively studied. The use of FPs in mixed models is a very attractive alternative to explain the within-subjects' measurements dependence in models where there is non-linearity in the relationship between the response variable and continuous covariates. This characteristic occurs because the FPs offers a variety of non-linear functional forms for the continuous covariates in the average response, in which the family of the conventional polynomials and some asymmetric curves with asymptotes stand out. The incorporation of FPs into the structure of the mixed models has been investigated by several authors. However, there are no works about the following issues: the modeling of the fixed and random effects (mainly in the presence of several continuous and categorical covariates), the study of the influence of the FPs on the structure of the random effects, the investigation of an adequate

structure for the covariance of the random errors, or, a point that has central importance to the selection of the model, to perform a diagnostic analysis of the fitted models. In our point of view, a contribution of great relevance is the investigation and the proposition of strategies for fitting FPs with mixed effects encompassing the points mentioned above, with the goals of filling these gaps and to awaken the users to the great potential of mixed models, now even more extended, for the modeling of correlated data. In this thesis we propose a strategy for the use of FPs to model nonlinear trends in grouped data, such as repeated measurements, longitudinal and multilevel data. In this proposal, we allow the inclusion of interaction components between FP transformations and categorical variables through the use of an adaptation of the MFPI (Multivariable Fractional Polynomial Interaction) algorithm proposed by Royston & Sauerbrei (2004). We also model the covariance structure and use graphical tools for diagnostic analysis (Singer et al., 2017) in order to verify possible violations of the model assumptions, as well as to evaluate the effect of influential observations on the fitting. We conclude with an analysis of three practical examples in which fittings under the mixed models are compared. We use the R programming language as computational support.

Keywords: Curvature, functional form, fractional polynomial model, longitudinal data, variable selection.

1 INTRODUÇÃO

Os modelos de regressão linear são utilizados para descrever a distribuição de uma variável resposta, ou dependente, em função de uma ou mais variáveis regressoras, explicativas ou independentes. O termo linear é usado para indicar que o modelo é linear nos parâmetros das variáveis regressoras. Esses modelos têm numerosas aplicações e estão presentes em quase todas as áreas do conhecimento, incluindo Engenharia, Ciências Físicas e Químicas, Economia, Administração, Ciências Biológicas e Ciências Sociais (Montgomery et al., 2012). Os principais objetivos são identificar e estabelecer a relação entre a variável resposta e as variáveis regressoras, permitindo desta forma, fazer previsões.

Variáveis regressoras contínuas são amplamente utilizadas em modelos de regressão. Na maioria das aplicações, tais variáveis são inseridas sem transformação e modeladas como tendo relações lineares com os valores esperados da resposta. No entanto, as variáveis regressoras podem contribuir de maneira não linear e a suposição de linearidade imposta não é apropriada.

Na pesquisa médica, caso a linearidade não seja encontrada, as variáveis contínuas são frequentemente convertidas em variáveis categóricas pelo agrupamento de valores em duas ou mais categorias. A categorização introduz problemas na definição do ponto de corte ótimo, no número de grupos, na superparametrização, na perda de informação e, conseqüentemente, no poder do teste (Royston & Sauerbrei, 2008). Embora a categorização de dados contínuos ainda seja popular, Royston et al. (2006) acreditam ser desnecessária para a análise estatística.

Do ponto de vista metodológico, as desvantagens no agrupamento de uma regressora são consideradas em muitos trabalhos, incluindo Altman et al. (1994), Mac-

Callum et al. (2002), Irwin & McClelland (2003), Austin & Brunner (2004) e Royston et al. (2006). Em geral, uma regressora contínua com ponto de corte pode não descrever de maneira apropriada um relacionamento curvilíneo entre uma variável regressora e uma variável resposta (Royston & Sauerbrei, 2008).

Outra abordagem utilizada na modelagem de relacionamentos curvilíneos entre a variável resposta e as variáveis regressoras é a quantificação dos efeitos das variáveis regressoras por meio do modelo de regressão polinomial. Há muito se reconhece que a regressão polinomial tem algumas fraquezas, em que se destacam a falta de flexibilidade em modelos de baixa ordem, como os quadráticos; a propensão a não ter um bom ajuste na presença de valores extremos em modelos de ordem mais alta; além da incapacidade de modelar relações que tenham assíntotas.

Certamente existem outros métodos que permitem modelar relações curvilíneas. A técnica de regressão não paramétrica, por exemplo, permite uma maior flexibilidade na possível forma da função, pois é necessário apenas escolher o espaço de funções apropriado, geralmente motivado pelo grau de suavidade que a função de regressão pode ter. Porém, esta grande vantagem tem um preço, estimadores não paramétricos são menos eficientes que os estimadores paramétricos quando o modelo paramétrico é válido (Dias, 2001).

Qualquer técnica estatística utilizada na construção de um modelo de regressão precisa atender a dois desafios primários: selecionar as variáveis regressoras que estão relacionadas à variável resposta de interesse e determinar a forma funcional apropriada para as variáveis regressoras contínuas. A construção de um modelo no qual as variáveis regressoras influentes e as relações funcionais são de interesse central requer uma classe mais ampla de formas funcionais (Sauerbrei et al., 2007a).

Uma abordagem alternativa é manter a variável regressora contínua e permitir alguma forma de não linearidade. Box & Tidwell (1962) sugeriram utilizar transformações potência nas variáveis regressoras contínuas. Uma família mais geral de modelos paramétricos, associada à abordagem de Box & Tidwell (1962), baseada em funções polinômios fracionários (FPs - *Fractional Polynomials*) é proposta por

Royston & Altman (1994).

Royston & Altman (1994) tratam principalmente de FPs para o caso de uma única variável regressora contínua no modelo, mas também sugerem e ilustram um algoritmo para ajustar FPs em modelos com várias variáveis regressoras. Esse algoritmo é denominado de procedimento MFP ou, simplesmente, MFP (*Multivariable Fractional Polynomials*). Uma abordagem da construção do MFP é descrita por Sauerbrei & Royston (1999) considerando várias variáveis regressoras contínuas e categóricas.

Royston & Sauerbrei (2004) estenderam o procedimento MFP para investigar possíveis interações entre variáveis regressoras contínuas e categóricas, propondo um novo procedimento, denominado MFPI (*MFP Interaction*), para o qual foi desenvolvido um algoritmo.

Os FPs são adequados para modelar dados não lineares e têm sido usados em muitas aplicações, tais como análise de regressão simples e múltipla, modelos lineares generalizados, análise de sobrevivência, análise de meta-regressão, meta-análise de curvas de sobrevivência, como mostram Sauerbrei & Royston (1999), Geys et al. (1999), Bossard et al. (2003), Berger et al. (2003), Faes et al. (2003), Bagnardi et al. (2004), Lambert et al. (2005), Shkedy et al. (2006), Faes et al. (2006), Sauerbrei et al. (2007b), Faes et al. (2007), Bollaerts et al. (2008), Jansen (2011), Wong et al. (2011), Jansen & Cope (2012), Baneshi et al. (2013), Austin et al. (2014) e Zhang (2016), Kasenda et al. (2016), Staley & Burgess (2017), entre outros. Royston & Sauerbrei (2008) consideram que, embora haja importantes diferenças teóricas e práticas entre esses modelos de regressão, o uso dos FPs dentro dessas estruturas introduz poucas complicações ou problemas novos.

A incorporação dos FPs na estrutura dos modelos mistos tem sido investigada por diversos autores. Por exemplo, no The European Collaborative Study (2003) os autores ajustaram um modelo misto usando FPs para a variável regressora idade e consideraram a modelagem da heterogeneidade de variâncias ao longo da idade e a correlação serial dos erros. No entanto, o trabalho não detalha a metodologia, a

complexidade envolvida na modelagem da estrutura da matriz de covariâncias e nem apresenta o modelo final ajustado. Long & Ryou (2010), Tan et al. (2011), Wen et al. (2012), Johnson et al. (2013) e Aregay et al. (2013) discutem FPs no contexto dos modelos mistos em que o comportamento da variável resposta foi explicado exclusivamente por transformações e funções FP para uma única variável regressora contínua. Utilizando também uma única variável regressora contínua Vanderstichel et al. (2013) e Tilling et al. (2014) exploram os FPs incluindo termos de interação entre as transformações FP selecionadas e uma variável regressora binária.

Embora o uso dos FPs em modelos mistos tenha recebido considerável atenção na literatura, vemos lacunas a serem preenchidas, tais como: a exploração da problemática da modelagem na parte fixa e na parte aleatória, principalmente na presença de várias variáveis regressoras contínuas e categóricas, o estudo da influência dos FPs na estrutura dos efeitos aleatórios, a investigação de uma adequada estrutura para a matriz de covariâncias para os erros e , um ponto de fundamental importância para colaborar com a seleção do modelo, a realização da análise de diagnóstico dos modelos ajustados.

Uma contribuição, do nosso ponto de vista, de grande relevância, é a investigação e oferecimento de estratégias para a implementação do uso de polinômios fracionários em modelos mistos, doravante denominado de modelos PFMMs, com o objetivo de preencher essas lacunas e de despertar os usuários para o imenso potencial dos modelos mistos, agora ainda mais ampliado, para a modelagem de dados correlacionados.

1.1 Objetivos

A presente tese tem como objetivo principal propor uma estratégia para a implementação do uso de polinômios fracionários em modelos mistos. Com vistas ao problema geral de várias variáveis regressoras contínuas, comum em dados observacionais, a estratégia explorada na tese segue de perto os procedimentos MFP e MFPI propostos por Royston & Altman (1994) e Royston & Sauerbrei (2004). Nesse

contexto, os objetivos específicos são:

- (1) construir o modelo polinômio fracionário com efeitos mistos (denominado PFMM);
- (2) apresentar uma estratégia para o ajuste do modelo PFMM;
- (3) construir funções na linguagem R (R Core Team, 2018) para implementação dos polinômios fracionários em modelos mistos;
- (4) aplicar o modelo PFMM a três conjuntos de dados que diferem em relação ao tipo e quantidade de variáveis regressoras, sendo dois com estrutura longitudinal e um hierárquico;
- (5) realizar análise de diagnóstico no modelo PFMM para cada uma das aplicações.

1.2 Organização da tese

Esta tese encontra-se dividida em cinco capítulos. Nos capítulos 2 e 3 apresentamos a revisão de literatura abordando os dois temas principais da tese que são o modelo de regressão polinômio fracionário e o modelo linear de efeitos mistos. No Capítulo 4 propomos o modelo polinomial fracionário no contexto dos modelos mistos e uma estratégia para análise do modelo proposto. Análises de três exemplos práticos são apresentadas no Capítulo 5. No Capítulo 6 concluímos com uma breve discussão e apresentamos algumas perspectivas de trabalhos futuros.

1.3 Suporte computacional

Os pacotes computacionais mais amplamente utilizados para o ajuste de modelos lineares mistos são `proc MIXED` do *software SAS* (Statistical Analysis Software), e os pacotes `lme4` e `nlme` da linguagem de programação R, de código aberto. Nesta tese é usada a linguagem de programação R (versão 3.5.1) que permite

não só o desenvolvimento de funções apropriadas para a análise dos diferentes modelos apresentados e o acesso aos dados de um dos exemplos analisados, mas também a modificação de funções já existentes, o que tem contribuído para sua popularização.

Usamos ainda alguns pacotes do R no desenvolvimento da tese. Em ordem alfabética, os pacotes usados e suas citações são as seguintes: `nlme` (Pinheiro et al., 2018), `lattice` (Sarkar, 2008), `mfp` (original de Gareth Ambler & modificado por Axel Benner, 2015), `Rmisc` (Hope, 2013), `ggplot2` (Wickham, 2016), `lmer` (Bates et al., 2015), `car` (Fox & Weisberg, 2011).

2 POLINÔMIOS FRACIONÁRIOS

2.1 Introdução

A maioria dos problemas que fazem uso de polinômios emprega apenas potências com valores inteiros nos termos das variáveis regressoras. Como uma alternativa, Royston & Altman (1994) propuseram os polinômios fracionários que são uma extensa família de modelos que consideram a transformação potência nas variáveis regressoras contínuas, em que os valores assumidos pelas potências são restritos a um conjunto predefinido de valores inteiros e não inteiros, proporcionando, desta forma, uma classe mais abrangente de funções médias e, mais ainda, usando apenas alguns termos com resultados semelhantes a outras famílias de transformações potência (Weisberg, 2005).

Os polinômios convencionais são um subconjunto da família dos FPs, assim como algumas curvas assimétricas e/ou com assíntotas. Os FPs têm sido extensivamente utilizados nos modelos de regressão linear múltipla com erros normais, nos modelos de Cox para dados censurados e nos modelos lineares generalizados (Royston & Sauerbrei, 2008). Mais recentemente, o uso dos FPs em modelos mistos recebeu considerável atenção na literatura por autores como Long & Ryou (2010), Tan et al. (2011), Wen et al. (2012), Johnson et al. (2013), Aregay et al. (2013), Vanderstichel et al. (2013) e Tilling et al. (2014).

Os FPs são extremamente úteis nos modelos de regressão por fornecerem uma parametrização flexível tendo sido empregados em uma grande variedade de cenários de pesquisa na área da saúde, como na realização de uma meta-análise de 34 estudos prospectivos para analisar a associação entre consumo de álcool e mortalidade total

(Di Castelnuovo et al., 2006); na área da Psicologia, como um meio de estimar trajetórias de crescimento não lineares com dados longitudinais (Long & Ryoo, 2010), na predição da imunidade a longo prazo de anticorpos anti-HPV, também utilizando dados longitudinais (Aregay et al., 2013) e para avaliar a relação entre dose cumulativa e exposição que é uma questão importante na epidemiologia, na saúde pública e na medicina clínica (Austin et al., 2014).

Além de pesquisas na área da saúde, os FPs foram aplicados em estudos de superfícies de resposta para a otimização de um biorreator (Gilmour & Trinca, 2005) e em vários estudos no campo da pecuária (Robert-Granié et al., 2002; Faes et al., 2006; Martins et al., 2011), entre outros.

Nesta seção introduzimos o modelo polinômio fracionário. Na Seção 2.2 distinguimos entre uma transformação FP e uma função ou modelo FP para uma única regressora contínua, apresentamos características das funções FPs e o procedimento para seleção da função FP. Na Seção 2.3 descrevemos a extensão do modelo FP para o caso com múltiplas variáveis regressoras denominado de procedimento MFP e, na Seção 2.4, explicitamos uma extensão do procedimento MFP disponível para modelar interações entre uma variável regressora contínua e uma variável categórica.

2.2 Polinômios fracionários para uma variável regressora contínua

Suponha que dados observados para as variáveis Y e x sugerem uma relação não linear implicando que o modelo de regressão linear simples $Y = \beta_0 + \beta_1 x + \epsilon$ não traduz a relação entre estas variáveis. Uma forma de contornar esse problema é considerar o preditor linear

$$\varphi_1(x, p) = \beta_0 + \beta_1 x^p, \quad (2.1)$$

em que a relação entre $\mathbb{E}(Y|x)$ e x^p é linear.

Em princípio, a transformação x^p em (2.1), com $x > 0$ e $p \in \mathbb{R}$, é definida como

uma transformação FP de grau 1 e denotada por FP1, e $\varphi_1(x, p)$ é definida como função ou modelo FP1. Inerente à definição da transformação FP é a condição de que $x > 0$. Se ocorrem valores não positivos em x , então é necessária uma transformação preliminar em x para garantir sua positividade (detalhes na Subseção 2.2.1).

Embora muitos valores diferentes de potências possam ser considerados nos modelos FPs, Royston & Altman (1994) propõem restringir a escolha da potência p ao conjunto $S = \{-2; -1; -0,5; 0; 0,5; 1; 2; 3\}$, com $p = 0$ indicando o logaritmo natural de x , isto é, $x^0 = \ln(x)$, argumentando que do ponto de vista prático esses valores fornecem flexibilidade suficiente para modelar as relações na maioria das aplicações.

No conjunto S são consideradas a não transformação ($p = 1$) e as transformações inversa ($p = -1$), logarítmica ($p = 0$), raiz quadrada ($p = 0,5$), quadrática ($p = 2$) e cúbica ($p = 3$).

A classe dos FPs para uma variável regressora descreve uma transformação potência na variável regressora contínua x , com $x > 0$, enquanto são possíveis diferentes graus para o FP. O grau do FP é determinado pelo número de potências na função.

Considerando o preditor linear

$$\varphi_2(x, \mathbf{p}) = \beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2} = \beta_0 + x^{\mathbf{p}} \boldsymbol{\beta}, \quad (2.2)$$

em que $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, $\mathbf{p} = (p_1, p_2)^\top \in S^2$ e $x > 0$, a transformação FP em (2.2) é definida como uma transformação FP de grau 2, denotada por FP2, e $\varphi_2(x, \mathbf{p})$ como função ou modelo FP2. Uma particularidade da transformação FP2 se dá na composição do vetor $x^{\mathbf{p}}$ quando temos potências distintas, $p_1 \neq p_2$, e quando temos potências repetidas, $p_1 = p_2$, sendo expresso por (Royston & Sauerbrei, 2008)

$$x^{\mathbf{p}} = x^{(p_1, p_2)} = \begin{cases} (x^{p_1}, x^{p_2}), & \text{se } p_1 \neq p_2 \\ (x^{p_1}, x^{p_1} \ln x), & \text{se } p_1 = p_2, \end{cases} \quad (2.3)$$

em que $x > 0$ e $\mathbf{p} \in S^2$. Portanto, para potências repetidas, $p_1 = p_2$, a função FP2

é dada por

$$\varphi_2(x, \mathbf{p}) = \beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_1} \ln x. \quad (2.4)$$

A Tabela 2.1 apresenta quatro exemplos de funções FP2, destacando a variedade de formas funcionais que os FPs podem assumir. Os Exemplos 1 e 2 ilustram o que ocorre se pelo menos uma potência assumir o valor zero. Os Exemplos 1 e 3 ilustram o caso de potências repetidas ($p_1 = p_2$), enquanto o Exemplo 4 ilustra que polinômios convencionais são um caso especial dos FPs.

Tabela 2.1. Exemplos ilustrativos de funções FP2 para uma variável regressora

Exemplo	$(p_1; p_2)$	Função
1	(0; 0)	$\beta_0 + \beta_1 \ln x + \beta_2 \ln x \ln x$
2	(0; -0,5)	$\beta_0 + \beta_1 \ln x + \beta_2 \left(\frac{1}{\sqrt{x}} \right)$
3	(-2; -2)	$\beta_0 + \beta_1 x^{-2} + \beta_2 x^{-2} \ln x$
4	(1; 2)	$\beta_0 + \beta_1 x + \beta_2 x^2$

Para potências distintas, $(p_1; p_2)$ é equivalente a $(p_2; p_1)$. Assim, com o conjunto S de potências, existem 8 transformações FP1 e 36 transformações FP2, sendo 28 com potências distintas e 8 com potências repetidas.

Além disso, Royston & Altman (1994) estendem os FPs para transformações com graus mais elevados, como FP3, FP4, etc., podendo ser útil ou, às vezes, até mesmo necessário para descrever os dados adequadamente (Mayer et al., 2016). Usando a relação recursiva, uma função FP de grau m , FP_m , para uma única variável regressora contínua x , $x > 0$, é dada por

$$\varphi_m(x, \mathbf{p}) = \beta_0 + \beta_1 x^{p_1} + \beta_1 x^{p_2} + \dots + \beta_m x^{p_m} = \sum_{r=0}^m \beta_r H_r(x) \quad (2.5)$$

em que m é um inteiro positivo, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_m)^\top$ é o vetor dos coeficientes de regressão e $H_r(x)$, com $H_0(x) = 1$ produzindo o intercepto, é a função que associa

a variável regressora contínua x à transformação FP definida por

$$H_r(x) = \begin{cases} x^{p_r}, & \text{se } p_r \neq p_{r-1} \\ H_{r-1}(x) \ln x, & \text{se } p_r = p_{r-1}, \end{cases} \quad (2.6)$$

com $r = 1, \dots, m$, $p_0 = 0$ e $\mathbf{p} = (p_1, p_2, \dots, p_m)^\top$, com $p_1 \leq \dots \leq p_m$, o vetor de potências.

Por exemplo, para $m = 3$ e $\mathbf{p} = (0; 2; 2)$, temos $H_1(x) = \ln x$, $H_2(x) = x^2$, $H_3(x) = x^2 \ln x$.

Royston & Altman (1994) consideram que o uso de modelos de grau $m \leq 2$ com potências selecionadas do conjunto S são suficientemente flexíveis para cobrir satisfatoriamente a maioria dos casos práticos. Baseados em estudos de simulação, Royston & Sauerbrei (2008) comprovam que mesmo relações muito complexas entre a variável resposta e uma ou mais variáveis regressoras podem ser adequadamente representadas por um FP2 com o vetor de potências $\mathbf{p} \in S^2$ e elencam algumas razões pelas quais não é atribuído a S o conjunto dos números reais:

1. a estimação é muito facilitada usando um conjunto menor;
2. para as funções FP2, por exemplo, apenas 36 modelos precisam ser ajustados para encontrar os melhores expoentes em S ;
3. uma classe restrita é especialmente importante na modelagem com múltiplas variáveis regressoras;
4. a estimação de potências requer otimização não linear e esses algoritmos nem sempre convergem;
5. as potências selecionadas de S podem ser vistas como uma aproximação da “melhor” escolha possível de potências no intervalo $[-2, 3]$, tanto para modelos FP1 como para FP2.

Um modelo FP de grau m tem $2m$ parâmetros no preditor linear, cada um utilizando um grau de liberdade (excluindo a constante β_0).

Uma vez assumida uma distribuição de probabilidade para Y , os parâmetros do modelo FP são estimados pelo método de máxima verossimilhança. Para cada possível potência $p \in S$, estimada por máxima verossimilhança aplicada ao conjunto S discreto, a estimação dos coeficientes de regressão consiste em encontrar valores para o vetor β que maximizam a verossimilhança dos modelos ajustados, conforme especificado na equação (2.5).

Para a classe FP1, oito modelos serão ajustados, enquanto para FP2 serão 36 modelos. Portanto, entre os oito modelos FP1, o melhor ajuste será aquele cuja potência p resultar na maior verossimilhança. Similarmente, para os 36 modelos FP2, aquele com melhor ajuste terá a maior verossimilhança para as potências p_1 e p_2 . A seleção da função FP, FP1 ou FP2, mais apropriada para a inclusão em um modelo de regressão será discutida na Subseção 2.2.2.

Em geral, as funções FP1 são monótonas e as funções FP2 são monótonas ou unimodais (Royston & Sauerbrei, 2008). Funções monótonas têm uma inclinação positiva ou negativa, enquanto funções unimodais têm um mínimo ou um máximo para algum valor positivo de x .

Para $p_1 \neq p_2$, a função FP2 será

$$\begin{cases} \text{monótona,} & \text{se } \text{ sinal}(\beta_1 \times \beta_2) \times \text{ sinal}(p_2) = \text{ sinal}(p_1), \text{ com } \text{ sinal}(0) = +1, \\ \text{unimodal,} & \text{caso contrário.} \end{cases}$$

Para $p_1 = p_2$, a função FP2 será sempre unimodal.

Quando uma função FP2 qualquer é unimodal, o valor de x em que ocorre o ponto de mínimo ou o de máximo depende dos valores assumidos pelas potências p_1 e p_2 e pelos coeficientes β_1 e β_2 . O valor de x referente ao mínimo ou máximo de uma função FP2 unimodal em termos de suas potências e seus coeficientes estão resumidos na Tabela 2.2 adaptada de Royston & Sauerbrei (2008).

Tabela 2.2. Ponto de mínimo ou máximo em x em uma função FP2 unimodal dadas suas potências e seus coeficientes ($\kappa = -\beta_1/\beta_2$)

Valor de p_2	Valor de p_1	
	$p_1 = 0$	$p_1 \neq 0$
$p_2 = 0$ ($p_2 \neq p_1$)	—	$(\kappa p_1)^{-1/p_1}$
$p_2 \neq 0$ ($p_2 \neq p_1$)	$(\kappa/p_2)^{1/p_2}$	$(\kappa p_1/p_2)^{1/(p_2 - p_1)}$
$p_2 = p_1$	$\exp(\kappa/2)$	$\exp(\kappa - 1/p_1)$

Para $x \rightarrow \infty$, as funções FP1 e FP2 possuem assíntotas quando $p < 0$ e quando ambos p_1 e p_2 forem negativos, respectivamente.

Embora bastante simples, a família FP1 é frequentemente útil. As Figuras 2.1 e 2.2 apresentam, esquematicamente, as oito formas das curvas FP1 disponíveis no conjunto S , separadas em potências negativas e positivas, respectivamente.

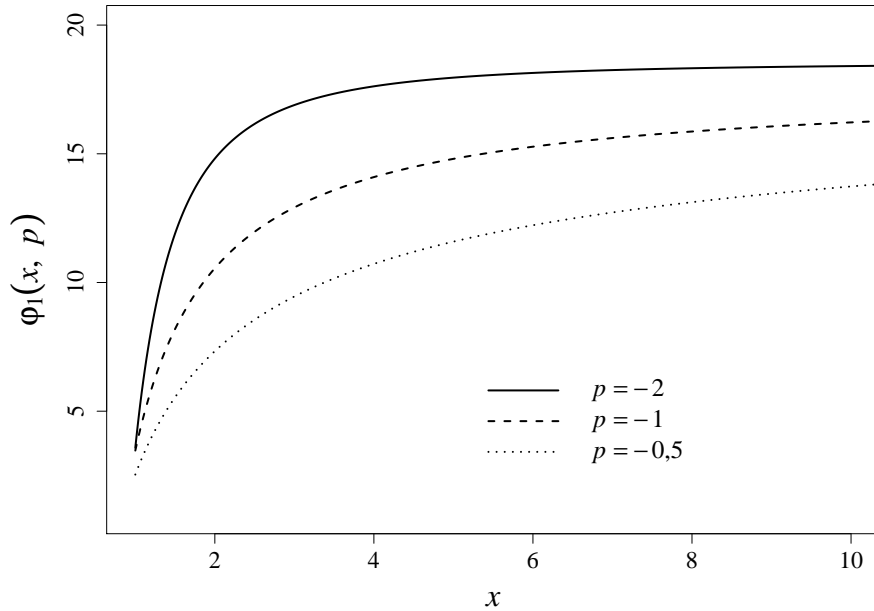


Figura 2.1: Ilustração das três curvas FP1 com potências $p < 0$ e coeficiente $\beta_1 < 0$.

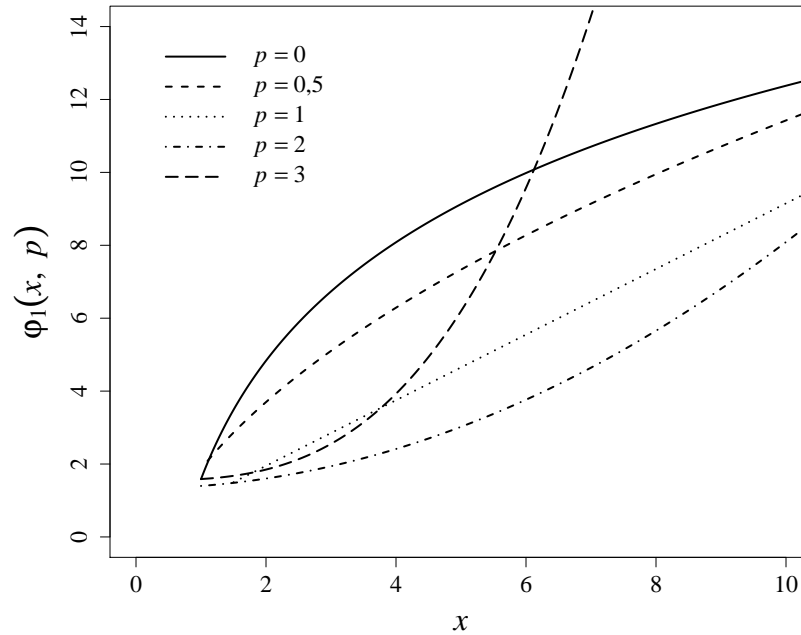


Figura 2.2: Ilustração das cinco curvas FP1 com potências $p \geq 0$ e coeficiente $\beta_1 > 0$.

A capacidade de gerar uma variedade de curvas, algumas das quais podem apresentar um crescimento ou decréscimo acentuados ou ainda uma parte quase plana e, como destacado, ter assíntotas, é uma característica particularmente útil das funções FP2 (Royston & Sauerbrei, 2008). A Figura 2.3 apresenta exemplos de seis formas de funções FP2, adaptadas de Long & Ryoo (2010), com algumas combinações de p_1 e p_2 e β_1 e β_2 , escolhidas para ilustrar a flexibilidade das funções FP2.

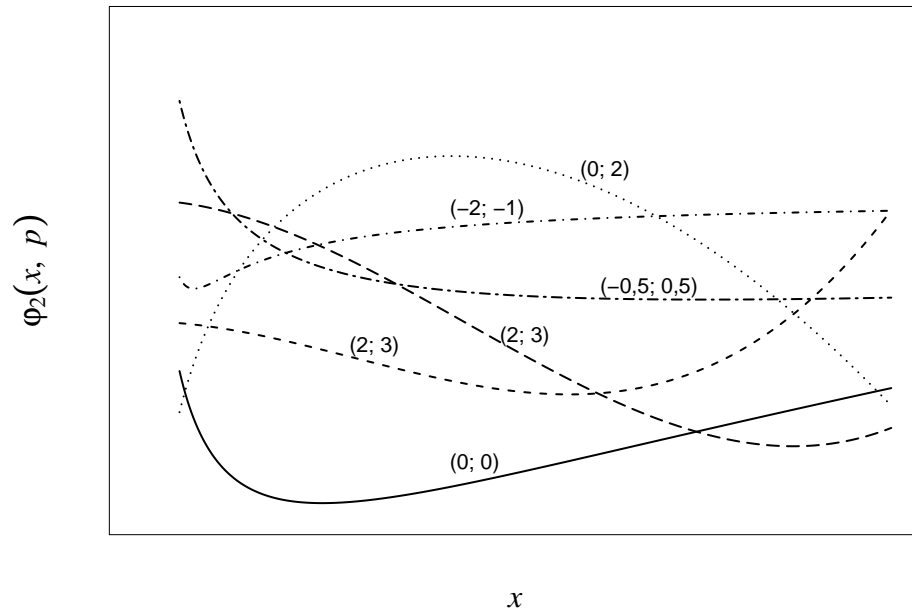


Figura 2.3: Ilustração de seis curvas FP2 com diferentes potências (p_1, p_2) e coeficientes (β_1, β_2) . Números entre parenteses indicam as potências.

2.2.1 Mudança da origem da regressora

O método original FP exige que todos os valores de uma variável regressora contínua sejam positivos. Entretanto, em conjuntos de dados com várias variáveis regressoras contínuas podemos ter algumas com valores de $x \leq 0$. Uma mudança de origem simples foi dada por Royston & Altman (1994) para lidar com regressoras nestas condições, escolhendo uma origem γ , por meio da seguinte expressão

$$x^* = x - x_{\min} + \gamma,$$

em que x_{\min} é o menor valor observado para x e considerando como possível escolha para γ o incremento mínimo entre os sucessivos valores amostrais ordenados de x . Se x for contagem a sugestão é utilizar $\gamma = 1$.

Uma correção alternativa foi proposta por Royston & Sauerbrei (2007) com o objetivo de fornecer uma abordagem formal para o problema de origem da variável regressora contínua e, ainda, reduzir a influência de observações extremas sobre a

seleção de modelos e sobre as estimativas dos seus parâmetros, evitando a remoção de observações do conjunto de dados. Os autores conduziram amplo estudo de simulação e, baseados nos resultados, sugeriram a seguinte transformação linear de x para o intervalo $[\delta, 1)$:

$$\omega_\delta(x) = \delta + (1 - \delta) \frac{(x - x_{min})}{(x_{max} - x_{min})}, \quad (2.7)$$

em que x_{max} é o maior valor observado para x e $\delta = 0,2$.

Existem outras propostas de correção porém estas são as mais simples e de maior apelo prático.

2.2.2 Procedimento de seleção da função

Ambler & Royston (2001) propõem a utilização de um procedimento de teste fechado para seleção da função, doravante denominado de FSP (*Function Selection Procedure*), que, se possível, simplifica um modelo FP para um de menor complexidade por meio de testes estatísticos apropriados. Um procedimento de teste fechado é um processo que permite a realização de vários testes de hipóteses simultaneamente controlando a probabilidade do Erro Tipo I.

O FSP proposto em 2001 é o procedimento padrão no softwares STATA, SAS e R (Sauerbrei et al., 2006), sendo uma versão melhorada do algoritmo de seleção de modelos proposto por Royston & Altman (1994).

O procedimento é conservador rejeitando a hipótese nula quando ela é verdadeira com probabilidade menor do que o nível de significância nominal. Os passos do FSP para seleção de modelos, cuja complexidade máxima aceita é para uma função FP2, é como segue:

1. Escolha a probabilidade de significância nominal α .
2. Teste o melhor modelo FP2 ($\beta_0 + \beta_1 x^{p1} + \beta_2 x^{p2}$) com nível α contra o modelo nulo (somente β_0) usando 4 graus de liberdade. Se o resultado do teste é não significativo, pare, concluindo que o efeito de x é não significativo com nível α . Caso contrário, continue.

3. Teste o melhor modelo FP2 contra o modelo linear $(\beta_0 + \beta_1 x)$ com nível α usando 3 graus de liberdade. Se o resultado do teste é não significativo, pare, e o modelo final é uma reta. Caso contrário, continue.
4. Teste o melhor modelo FP2 contra o melhor modelo FP1 $(\beta_0 + \beta_1 x^p)$ com nível α usando 2 graus de liberdade. Se o resultado do teste é não significativo, pare, e o modelo final é do tipo FP1. Caso contrário, o modelo final é FP2. Fim do procedimento.

O teste, no passo 2, é um teste de associação global da resposta com x . No passo 3, examina evidência de não linearidade e, no passo 4, escolhe entre um modelo não linear simples ou mais complexo. Em geral, escolhe-se $\alpha = 0,05$.

As hipóteses a serem testadas em todos os passos do FSP se referem aos efeitos fixos dos modelos, cuja estimação é feita por máxima verossimilhança, e os testes realizados são baseados na razão de verossimilhanças generalizada (RVG). Considerando LL_1 o máximo da log-verossimilhança sob o modelo com menor número de parâmetros e LL_2 o máximo da log-verossimilhança sob o modelo com maior número de parâmetros, a estatística de teste da RVG é dada por

$$T_{RVG} = -2(LL_1 - LL_2). \quad (2.8)$$

Como o interesse está na comparação de modelos encaixados que diferem na especificação de seus efeitos fixos, a distribuição assintótica de (2.8), sob a hipótese nula de que o modelo com menor número de parâmetros é adequado, tem distribuição χ_r^2 , com r sendo a diferença de parâmetros entre os modelos com o maior e o menor número de parâmetros.

Um fato importante a ser destacado é quanto à utilização da função componente do desvio (*Deviance*) nos modelos FPs. Um modelo com o máximo da log-verossimilhança é equivalente a um modelo com a menor componente do desvio, D , em que $D = -2LL$, com LL representando o máximo da log-verossimilhança do modelo em questão. Considerando que a literatura sobre FPs apresenta referência

ao uso da componente do desvio, fizemos uso desta componente para explicar como modelos FP1 estão encaixados nos modelos FP2, de forma análoga ao descrito em Royston & Sauerbrei (2008).

As comparações entre todos os modelos no algoritmo FSP são realizadas pelo teste da razão de verossimilhanças generalizado. Royston & Sauerbrei (2008) utilizam o seguinte argumento para mostrar que modelos FP1 estão encaixados nos modelos FP2. Seja D_1^* a componente do desvio do melhor modelo FP1 com potência p_1^* , D_2^* a componente do desvio do melhor modelo FP2 com potências (p_1, p_2) e D_2 a componente do desvio dos modelos FP2 com potências (p_1^*, p_2) , denotado por FP2- p_1^* . Na Tabela 2.3 apresentamos as relações entre as três componentes.

Tabela 2.3. Comparando modelos FP1 e FP2

	Modelo		
	FP1	FP2- p_1^*	FP2
Potências	p_1^*	(p_1^*, p_2)	(p_1, p_2)
Componente do desvio	D_1^*	$D_2 \leq D_1^*$	$D_2^* \leq D_2$

Para todo modelo FP1 com potência p_1^* existem oito modelos FP2 com potências (p_1^*, p_2) , com $p_2 \in S$, no qual o modelo FP1 está encaixado, ou seja, é um caso particular quando $\beta_2 = 0$. Desta forma, a componente do desvio, D_2 de qualquer modelo FP2 com potências (p_1^*, p_2) é menor ou igual a componente do desvio D_1^* do melhor modelo FP1. Lembrando que o melhor modelo FP2 tem sempre a menor componente do desvio, tem-se que $D_2^* \leq D_2$. Assim,

$$D_2^* \leq D_2 \leq D_1^*$$

e, portanto, a diferença de componente do desvio entre o melhor modelo FP1 e o melhor modelo FP2, ou seja, $D_1^* - D_2^*$, é sempre não negativa.

2.3 Polinômios fracionários para várias variáveis regressoras

O procedimento MFP, ou simplesmente MFP, permite que funções FP para várias variáveis regressoras sejam estimadas simultaneamente, sendo uma extensão dos modelos FPs univariados. O MFP combina a seleção de variáveis com a determinação da forma funcional para as variáveis regressoras contínuas, examinando, de forma sistemática, se o efeito de cada variável regressora contínua é melhor modelado por uma função não linear por meio da classe de funções FP ou por uma função linear.

Royston & Sauerbrei (2008) definem que o principal parâmetro requerido pelo MFP é o nível de significância nominal, sendo necessários dois níveis de significância, α_1 para a seleção das variáveis, realizado pelo método de eliminação *backward*, e α_2 para a comparação dos modelos FPs, realizado pelo procedimento FSP descrito na Subseção 2.2.2. A notação $\text{MFP}(\alpha_1, \alpha_2)$ indica um modelo selecionado utilizando α_1 para o processo de eliminação *backward*, BE (*Backward Elimination*), e α_2 para o procedimento FSP. Para o caso de $\alpha_1 = \alpha_2 = \alpha$ a notação será representada por $\text{MFP}(\alpha)$.

O algoritmo MFP permite que se especifique a complexidade máxima (grau) da função FP, além de poder eliminar variáveis regressoras não influentes por meio do método BE com nível de significância predefinido. O algoritmo procede em ciclos. A significância e a forma funcional de cada variável regressora contínua, por sua vez, são determinadas de forma univariada, ajustando-se juntamente com todas as variáveis regressoras contínuas, binárias e categóricas, presentes no modelo. A convergência ocorre quando não há mais alterações nas variáveis selecionadas e nem nas suas transformações FP.

2.3.1 Algoritmo MFP

Seja L o número de variáveis regressoras com potencial de inclusão no modelo. O algoritmo MFP, conforme implementação no STATA e no pacote `mfp` do R, segue

as seguintes etapas:

1. Fixe os níveis de significância α_1 e α_2 para seleção das variáveis regressoras e determinação da forma funcional das variáveis regressoras contínuas, respectivamente. Valores típicos são $\alpha_1 = \alpha_2 = 0,05$, porém estas podem diferir entre as variáveis regressoras. Fixando $\alpha_1 = 1$ para uma determinada variável regressora força sua inclusão no modelo. Fixando $\alpha_2 = 1$ para uma variável regressora contínua força o uso da função mais complexa para esta variável regressora, não havendo seleção da função.
2. Escolha o grau de liberdade máximo permitido para as funções FP, em que 4, 2 e 1 são os graus de liberdade máximo permitidos para um modelo FP2, FP1 e linear, respectivamente. A sugestão é utilizar o valor padrão de 4 graus de liberdade.
3. Ajuste um modelo linear usual com todas as variáveis regressoras a fim de determinar a ordem em que elas serão visitadas no procedimento MFP. As variáveis regressoras são ordenadas pelas probabilidades de significância apresentadas no modelo linear usual ajustado. A variável regressora mais significativa é visitada por primeiro no procedimento MFP e a menos significativa por último. Essa ordenação é mantida em todos os ciclos do procedimento.
4. Faça $c = 0$, para iniciar a contagem dos ciclos.
5. Faça $j = 1$ para iniciar a contagem das variáveis regressoras dentro de cada ciclo.
6. Se a j -ésima variável regressora, x_j , é contínua vá para o passo 7. Caso contrário, x_j é categórica ou binária e sua significância é testada no nível α_1 no ajuste com todas as outras variáveis regressoras incluídas. Se x_j é significativa, então será retida; caso contrário, será descartada. Vá para o passo 8.
7. O passo 2 do FSP é aplicado para x_j no nível $\alpha = \alpha_1$ (veja Subseção 2.2.2). Se x_j não for significativa será descartada. Caso contrário, os passos 3 e 4 do

FSP são aplicados no nível $\alpha = \alpha_2$ para selecionar uma função FP ou linear, no ajuste com todas as outras variáveis regressoras incluídas. Sendo selecionado um FP1 com potência p_1 ou FP2 com potências \mathbf{p} , x_j será substituída por $x_j^{p_1}$ ou $x_j^{\mathbf{p}}$ nas etapas subsequentes nas quais as outras variáveis regressoras serão visitadas.

8. Incluindo ou descartando x_j ela será reavaliada no próximo ciclo.
9. Faça $j = j + 1$. Se $j \leq L$, retorne ao passo 6 para processar a próxima variável regressora. Caso contrário, o c -ésimo ciclo está completo. Vá para o passo 10.
10. Faça $c = c + 1$. Se $c \geq c_{max}$, pare e informe que o algoritmo não convergiu em c_{max} ciclos. Uma sugestão é fixar $c_{max} = 5$. Caso contrário, vá para o passo 12.
11. Verifique se o conjunto de variáveis regressoras incluídas ou as transformações FP mudaram do ciclo $c - 1$ para o ciclo c . Em caso afirmativo, vá para o passo 5 para iniciar um novo ciclo. O novo ciclo começa com o melhor modelo ajustado no ciclo anterior em vez do modelo linear completo e cada variável regressora é considerada novamente para inclusão ou exclusão. Caso contrário, pare e informe as estimativas do modelo atual. Fim do procedimento.

Royston & Sauerbrei (2008) consideram que a convergência é geralmente alcançada em, no máximo, 4 ciclos.

2.4 Extensão do procedimento MFP para investigar interações

A investigação sobre a necessidade de inclusão de termos de interação entre variáveis regressoras é importante nas aplicações. No contexto de FPs, Royston & Sauerbrei (2004) propõem abordagens para os casos de interação entre uma variável contínua e uma variável categórica, como também para interação entre duas variáveis contínuas. Forneceremos a descrição da abordagem para o caso de uma variável

regressora categórica. Por simplicidade de notação consideramos que a variável regressora categórica é binária, porém o procedimento se estende facilmente para mais de duas categorias.

Considere x uma variável contínua e w uma variável binária assumindo valores $\{0, 1\}$. A fim de usar toda a informação da variável regressora contínua, mas ao mesmo tempo permitir uma possível não linearidade em x em todos os níveis de w , Royston & Sauerbrei (2004) propuseram o algoritmo MFPI com a finalidade de investigar interações entre x e w . O algoritmo é uma extensão do procedimento MFP para a seleção simultânea de variáveis e da forma funcional para uma variável regressora contínua.

2.4.1 Algoritmo MFPI

Com o algoritmo MFPI um modelo para x e w é determinado encontrando a melhor transformação FP para x e formando termos de interação multiplicativos com os níveis de w , cujas significâncias são avaliadas. Seguindo Royston & Sauerbrei (2008) o procedimento é realizado como segue:

1. Encontre, pelo método de máxima verossimilhança, o melhor modelo FP2, com potências $p_1, p_2 \in S$, para x , sempre incluindo w nos ajustes. Represente as transformações FP2 por $x^{\mathbf{P}} = (x^{p_1}, x^{p_2})$.
2. Para a potência $p_i, i = 1, 2$, defina novas variáveis regressoras: $x_{1i} = x^{p_i}w$, se $w = 1$, e $x_{0i} = 0$, caso contrário.
3. Realize o teste da razão de verossimilhanças generalizado para a interação $x \times w$ entre os modelos encaixados $w, x^{p_1}, x^{p_2}, x_{11}, x_{12}$ e w, x^{p_1}, x^{p_2} . A diferença entre as componentes do desvio é comparada com uma distribuição χ^2 com 2 graus de liberdade.

De acordo com Royston & Sauerbrei (2008) a razão para se ajustar funções FP2 em x , em vez de funções mais simples, é encontrar o melhor ajuste por meio de uma

classe flexível. Em termos da relação entre viés/variância, os autores argumentam que o aumento da variância decorrente do uso de potências FP2 para x pode ser tolerado dado o baixo valor do viés. Frisam que para evitar um superajustamento, causando sérios danos às funções ajustadas, a estimação de diferentes potências em cada nível de w não é considerada. Uma função FP2 com as mesmas potências em cada nível de w já é uma especificação bastante flexível.

O MFPI permite o ajuste agregando ao modelo outras variáveis regressoras influentes. Royston & Sauerbrei (2004) propõem ajustar um modelo em uma etapa preliminar, preferencialmente com o algoritmo MFP, sem considerar a variável regressora x . Para mais detalhes ver Royston & Sauerbrei (2004).

3 MODELOS LINEARES MISTOS

Dados correlacionados surgem com frequência em análises estatísticas e isto pode ser devido ao desenho amostral ou experimental, caracterizando padrões de agrupamento, tais como, levantamentos de dados sob sistema de amostragem complexo (alunos em salas de aula ou escolas, pacientes em hospitais, residentes em bairros, ratos em ninhadas, vacas em rebanhos, entre outros) ou ainda quando o conjunto de dados se enquadra nos estudos de medidas repetidas, isto é, quando se mede a mesma unidade amostral em mais de uma condição de avaliação (tempo, distância, peso, etc). É importante ressaltar que os estudos com dados longitudinais constituem um caso particular dos estudos com medidas repetidas, em que a condição de avaliação não pode ser aleatorizada, sendo o tempo a mais comum.

Quando o estudo envolve dados dessa natureza, as análises de variância clássicas ou de regressão, que são tradicionalmente adotadas, conduzem à violação da independência que estas técnicas pressupõem. Negligenciar as suposições adotadas pode levar a consequências prejudiciais para a inferência.

Nessas circunstâncias, os modelos mistos são a abordagem mais apropriada para a investigação de mudanças individuais ao longo das condições de avaliação e para o estudo dos efeitos dos fatores que podem influenciar a mudança. Dentre a classe de modelos mistos, o modelo linear misto tornou-se uma importante ferramenta estatística sendo utilizado nas mais variadas áreas do conhecimento, como ciências sociais, ciências biológicas e saúde (Diggle et al., 2002). A crescente popularidade dos modelos mistos pode ser explicada (1) pela versatilidade que oferecem na modelagem da correlação dentro da unidade amostral ou grupo; (2) pela possibilidade de serem utilizados em conjuntos de dados balanceados e não balanceados em uma estrutura

unificada; e (3) pela disponibilidade de softwares eficientes e confiáveis para ajustá-los (Pinheiro & Bates, 2000).

O modelo linear de efeitos mistos é um modelo linear que contém efeitos fixos e aleatórios, o que proporciona uma estrutura flexível para modelar dados correlacionados (Henderson, 1975; Laird & Ware, 1982). Em Henderson (1975) é apresentado o modelo de componentes de variâncias e a derivação das equações hoje conhecidas como equações de Henderson, que são úteis para, conjuntamente, estimar os efeitos fixos e prever os efeitos aleatórios. No início da década de 1980, Laird & Ware (1982), fazendo uso das ideias de Harville (1977), propuseram uma abordagem que utiliza os modelos mistos em dois estágios, possibilitando variações nas estruturas das matrizes de covariâncias das observações entre unidades e dentro das unidades amostrais.

A formalização estatística do modelo linear misto e resultados metodológicos relacionados são apresentados nas Seções 3.1 a 3.5.

3.1 Especificação do modelo

Considere N unidades amostrais ou grupos e n_i observações feitas na i -ésima unidade, $i = 1, \dots, N$, totalizando $n = \sum_{i=1}^N n_i$ observações. O modelo linear misto usual para a i -ésima unidade selecionada na amostra é

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, N, \quad (3.1)$$

em que $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^\top$ é o vetor $(n_i \times 1)$ de respostas da variável aleatória \mathbf{Y} , \mathbf{X}_i é a matriz de especificação conhecida de dimensão $(n_i \times p)$, associada aos parâmetros fixos, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é o vetor $(p \times 1)$ de parâmetros fixos, \mathbf{Z}_i é a matriz de especificação conhecida de dimensão $(n_i \times q)$, associada aos efeitos aleatórios, $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^\top$ é o vetor $(q \times 1)$ de efeitos aleatórios, não observáveis, com q representando o número de efeitos aleatórios e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{in_i})^\top$ é o vetor $(n_i \times 1)$ de erros aleatórios dentro da i -ésima unidade amostral. Admite-se, ainda,

que

$$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathcal{D}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathcal{R}_i), \quad \text{com } \mathbf{b}_i \perp \boldsymbol{\epsilon}_i, \quad (3.2)$$

em que \mathcal{D} e \mathcal{R}_i são matrizes simétricas positivas definidas de dimensões $(q \times q)$ e $(n_i \times n_i)$, respectivamente, e funções de parâmetros desconhecidos, tais que

$$\mathcal{D} = \sigma^2 \mathbf{D}(\boldsymbol{\theta}_D) \quad \text{e} \quad \mathcal{R}_i = \sigma^2 \mathbf{R}_i(\boldsymbol{\theta}_R), \quad (3.3)$$

com $\boldsymbol{\theta} = (\boldsymbol{\theta}_D^\top, \boldsymbol{\theta}_R^\top, \sigma^2)^\top$ sendo $\boldsymbol{\theta}_D$ e $\boldsymbol{\theta}_R$ os vetores de parâmetros de variâncias e covariâncias dos elementos de \mathbf{b}_i e $\boldsymbol{\epsilon}_i$, respectivamente, e σ^2 o parâmetro de dispersão desconhecido. Nas Seções 3.3 e 3.4 especificamos algumas estruturas para a matriz \mathbf{R}_i e descrevemos o teste para a seleção das componentes de variância da matriz \mathbf{D} , respectivamente.

Segue de (3.2) e (3.3) que \mathbf{Y}_i 's são vetores aleatórios independentes com distribuição normal multivariada com média $\mathbf{X}_i \boldsymbol{\beta}$ e matriz de covariâncias dada por

$$\mathbb{V}(\mathbf{Y}_i) = \mathbf{V}_i = \sigma^2 (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i). \quad (3.4)$$

De (3.1) a (3.3), segue que, condicionalmente aos efeitos aleatórios \mathbf{b}_i , \mathbf{Y}_i tem distribuição normal n_i variada, com média e variância definidas, respectivamente, por

$$\mathbb{E}(\mathbf{Y}_i | \mathbf{b}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i \quad (3.5)$$

$$\mathbb{V}(\mathbf{Y}_i | \mathbf{b}_i) = \sigma^2 \mathbf{R}_i, \quad (3.6)$$

com $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{i,n_i})^\top$ e

$$\mathbb{E}(y_{ij} | \mathbf{b}_i) = \mu_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i. \quad (3.7)$$

3.2 Estimação dos parâmetros

Considerando todas as N unidades amostrais, o modelo (3.1) pode ser reescrito da seguinte forma

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad (3.8)$$

que, matricialmente, é dado por

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{bmatrix},$$

Assumindo

$$\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \mathcal{N}_{(Nq+n)} \left(\begin{bmatrix} \mathbf{0}_{Nq} \\ \mathbf{0}_n \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix} \right), \quad (3.9)$$

temos que

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V}) \quad \text{com} \quad \mathbf{V} = \mathbf{V}_i \otimes \mathbf{I}_N = \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^\top + \boldsymbol{\Sigma}, \quad (3.10)$$

em que

$$\boldsymbol{\Delta} = \mathbf{I}_N \otimes \mathbf{D} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_N \end{bmatrix}. \quad (3.11)$$

com \mathbf{I}_N denotando a matriz identidade ($N \times N$) e \otimes o produto de Kronecker. Assim, os parâmetros do modelo (3.10) a serem estimados são os de efeitos fixos $\boldsymbol{\beta}$ e as componentes de covariância da matriz \mathbf{V} .

Para essa estimação, duas situações são possíveis: a primeira é que os valores das componentes de variância do vetor $\boldsymbol{\theta}$, das matrizes $\boldsymbol{\Delta}$ e $\boldsymbol{\Sigma}$, são conhecidos e a segunda, que é a que ocorre na prática, que são desconhecidos. Em ambos os

casos, a literatura é farta em mostrar métodos de estimação, tais como o de Mínimos Quadrados (MQ), Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR), apresentados por autores como Patterson & Thompson (1971), Robinson (1991), Searle et al. (1992), Draper & Smith (1998), Pinheiro & Bates (2000), Verbeke & Molenberghs (2000), McCulloch et al. (2008) e Demidenko (2013), por exemplo.

No primeiro caso, os estimadores obtidos através do método de MV são não viesados e resultam no melhor estimador linear não viesado (BLUE - *Best Linear Unbiased Estimator*) para β e no melhor preditor linear não viesado (BLUP - *Best Linear Unbiased Predictor*) para o vetor de efeitos aleatórios \mathbf{b} , dados, respectivamente, por

$$\tilde{\beta} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y} \quad (3.12)$$

e

$$\tilde{\mathbf{b}} = \Delta \mathbf{Z}^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \tilde{\beta}) = \Delta \mathbf{Z}^\top \mathbf{Q} \mathbf{Y}, \quad (3.13)$$

em que $\mathbf{Q} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}$, tal que $\mathbf{Q} \mathbf{X} = \mathbf{0}$ (propriedades de \mathbf{Q} são apresentadas em Nobre (2004)). O BLUE e o BLUP desempenham papéis importantes na análise e aplicação de modelos lineares mistos. Detalhes podem ser obtidos em Robinson (1991), por exemplo.

Na segunda situação, quando os componentes de variância θ são desconhecidos, a teoria mostra que a utilização do método de máxima verossimilhança produz estimadores viesados para θ . Neste caso, a utilização do método de máxima verossimilhança restrita, proposto por Patterson & Thompson (1971), resulta em estimadores, em geral, assintoticamente não viesados, digamos $\hat{\theta}$. Substituindo $\mathbf{V}(\hat{\theta}) = \hat{\mathbf{V}}$ em (3.12) e (3.13) resulta nas expressões, frequentemente referidas como BLUE empírico (EBLUE) e BLUP empírico (EBLUP), dadas, respectivamente, por

$$\hat{\beta} = (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{Y} \quad (3.14)$$

$$\hat{\mathbf{b}} = \hat{\Delta} \mathbf{Z}^\top \hat{\mathbf{V}}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}). \quad (3.15)$$

Jiang (1998, 1999) discute propriedades para o EBLUE e EBLUP.

3.3 Estrutura para a matriz de covariâncias \mathbf{R}_i

No contexto dos modelos mistos, como indicado em Singer et al. (2018), a covariância entre as observações obtidas em uma mesma unidade amostral poderá ser modelada direta ou indiretamente. Será diretamente se for por meio da matriz \mathbf{R}_i , que representa a estrutura de covariância das observações dentro das unidades amostrais. E, indiretamente, caso seja modelada por meio de uma combinação da matriz \mathbf{D} , que representa a variabilidade entre as unidades amostrais, com a matriz \mathbf{R}_i .

Para a modelagem da matriz de covariâncias dos efeitos aleatórios \mathbf{D} podemos iniciar utilizando a matriz positiva definida simétrica não estruturada. No entanto, em algumas aplicações práticas podemos restringir \mathbf{D} a formas especiais, a fim de que seja parametrizada por poucos parâmetros. Supondo, por exemplo, que os efeitos aleatórios sejam independentes, \mathbf{D} seria uma matriz diagonal, ou ainda que, além de serem independentes tenham a mesma variância, neste caso \mathbf{D} seria um múltiplo da matriz identidade.

Para a modelagem da matriz de covariâncias dos erros \mathbf{R}_i seguimos as ideias de Pinheiro & Bates (2000), que decompõem \mathbf{R}_i empregando conjuntos disjuntos de parâmetros para as matrizes \mathbf{A}_i e \mathbf{C}_i da seguinte forma

$$\mathbf{R}_i = \mathbf{A}_i \mathbf{C}_i \mathbf{A}_i, \quad (3.16)$$

em que \mathbf{A}_i é uma matriz diagonal com elementos não negativos, a fim de garantir a unicidade da matriz, e \mathbf{C}_i é uma matriz positiva definida com todos elementos da diagonal iguais a um.

Usando \mathbf{A}_i em (3.16), permitimos a heterocedasticidade das observações dentro da i -ésima unidade amostral, enquanto empregando a matriz de correlação \mathbf{C}_i , permitimos a correlação entre as observações dentro da unidade.

Mais especificamente tem-se que os elementos da matriz diagonal \mathbf{A}_i são expressos por

$$\{\mathbf{A}_i\}_{j,j} = g_{i,j,j} = g(\mu_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), \quad i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad (3.17)$$

com μ_{ij} definido em (3.7), \mathbf{v}_{ij} o vetor de variáveis regressoras conhecidas relacionadas com a variância (doravante denominada de variáveis regressoras da variância), definindo a função de variância para a j -ésima observação da unidade amostral i , $\boldsymbol{\delta}$ é o vetor dos parâmetros de variâncias, comum a todas as observações e $g(\cdot)$ a função de variância, assumida contínua em relação a $\boldsymbol{\delta}$. Maiores detalhes na Seção 3.3.1.

Além disso, considera-se que a matriz \mathbf{C}_i é especificada por um conjunto de parâmetros $\boldsymbol{\rho}$ que serão definidos na Seção 3.3.2. Desta forma, (3.16) ficará expressa como

$$\mathbf{R}_i(\mu_{ij}, \boldsymbol{\theta}_R, \mathbf{v}_{ij}) = \mathbf{A}_i(\mu_{ij}, \boldsymbol{\delta}, \mathbf{v}_{ij}) \mathbf{C}_i(\boldsymbol{\rho}) \mathbf{A}_i(\mu_{ij}, \boldsymbol{\delta}, \mathbf{v}_{ij}), \quad (3.18)$$

com $\boldsymbol{\theta}_R = (\boldsymbol{\delta}^\top, \boldsymbol{\rho}^\top)^\top$.

Pinheiro & Bates (2000) ressaltam que a decomposição de \mathbf{R}_i é conveniente tanto em termos teóricos como computacionais, pois permite modelar a estrutura de variância e de correlação separadamente e combiná-las em uma família flexível de modelos para a variância e covariância dentro da unidade amostral. Das possíveis estruturas disponíveis no pacote `nlme` (Pinheiro et al., 2018) da linguagem R (R Core Team, 2018) algumas estão descritas na Subseção 3.3.1.

3.3.1 Estrutura de variância

As funções de variância são utilizadas para modelar a estrutura de variância dos erros dentro da unidade amostral valendo-se de variáveis regressoras, valores ajustados ou de variáveis de estratificação. Pinheiro & Bates (2000), procedendo como em Davidian & Giltinan (1995), definem o modelo para a função de variância geral por

$$\mathbb{V}(\epsilon_{ij} | \mathbf{b}_i) = \sigma^2 g^2(\mu_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), \quad i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad (3.19)$$

com μ_{ij} , \mathbf{v}_{ij} , $\boldsymbol{\delta}$ e $g(\cdot)$ definidos em (3.17).

A formulação da função de variância (3.19) é muito flexível e intuitiva, porque permite que a variação dentro do grupo dependa dos efeitos fixos, $\boldsymbol{\beta}$, e dos efeitos aleatórios, \mathbf{b}_i , através dos valores esperados, μ_{ij} . No entanto, apresenta algumas dificuldades teóricas e computacionais, quando os erros dentro do grupo e os efeitos aleatórios não puderem ser considerados independentes. Por esse motivo, Pinheiro & Bates (2000), seguindo as sugestões de Davidian & Giltinan (1995), usam um modelo de variância aproximado em que os valores esperados μ_{ij} são substituídos por seus BLUPs $\hat{\mu}_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \hat{\mathbf{b}}_i$, tal que

$$\mathbb{V}(\epsilon_{ij}) \simeq \sigma^2 g^2(\hat{\mu}_{ij}, \mathbf{v}_{ij}, \boldsymbol{\delta}), \quad i = 1, \dots, N, \quad j = 1, \dots, n_i. \quad (3.20)$$

Caso o modelo de variância condicional (3.19) dependa de μ_{ij} , Pinheiro & Bates (2000) indicam um esquema reponderado iterativamente para o algoritmo de otimização.

O exemplo mais simples de uma função de variância é a $\text{varFixed}(v_{ij}) = \sqrt{v_{ij}}$, sendo utilizada quando a variância dentro da unidade é caracterizada por uma constante de proporcionalidade. Esta função atribui pesos determinados pelos valores da variável regressora da variância, dados por $w_{ij} = 1/v_{ij}$, $v_{ij} > 0$. A função $\text{varFixed}(\cdot)$ não depende de $\boldsymbol{\delta}$ nem de μ_{ij} e não utiliza variáveis de estratificação. Simplesmente implica num modelo com pesos conhecidos.

Com exceção da $\text{varFixed}(v_{ij})$, outras funções de variância permitem a inserção de uma ou várias variáveis de estratificação. Para refletir isso na notação considere-se que as observações podem ser divididas em vários estratos indexados por s , com $s = 1, \dots, S$. Exemplos de funções de variância são apresentados nas Tabelas 3.1, 3.2 e 3.3, adaptadas de Gałeczki & Burzykowski (2013).

Tabela 3.1. Exemplos de funções de variância que dependem de $\boldsymbol{\delta}$, mas não dependem de μ ^a

Função $g(\cdot)$	g_{ij}	Descrição
$\text{varPower}(\boldsymbol{\delta}, v_{ij}, s_{ij})$	$ v_{ij} ^{\delta_{s_{ij}}}$	potência de uma regressora da variância (v_{ij})
$\text{varExp}(\boldsymbol{\delta}, v_{ij}, s_{ij})$	$\exp(v_{ij}\delta_{s_{ij}})$	exponencial de uma regressora da variância (v_{ij})
$\text{varConstPower}(\boldsymbol{\delta}, v_{ij}, s_{ij})$	$\delta_{1,s_{ij}} + v_{ij} ^{\delta_{2,s_{ij}}}$	constante mais função de variância potência, $\delta_{1,s_{ij}} > 0$
$\text{varIdent}(\boldsymbol{\delta}, s_{ij})$	$\delta_{s_{ij}}$	diferentes variâncias por estrato, $\delta_1 = 1, \delta_s > 0$ para $s \neq 1$

^a Nome das funções usadas na primeira coluna corresponde ao nome usado no pacote nlme.

Tabela 3.2. Exemplos de funções de variância que dependem de $\boldsymbol{\delta}$ e μ ^a

Função $g(\cdot)$	g_{ij}	Descrição
$\text{varPower}(\boldsymbol{\delta}, \mu_{ij}, s_{ij})$	$ \mu_{ij} ^{\delta_{s_{ij}}}$	potência em $ \mu_{ij} $
$\text{varExp}(\boldsymbol{\delta}, \mu_{ij}, s_{ij})$	$\exp(\mu_{ij}\delta_{s_{ij}})$	exponencial em μ_{ij}
$\text{varConstPower}(\boldsymbol{\delta}, \mu_{ij}, s_{ij})$	$\delta_{1,s_{ij}} + \mu_{ij} ^{\delta_{2,s_{ij}}}$	constante mais função de variância potência, $\delta_{1,s_{ij}} > 0$

^a Nome das funções usadas na primeira coluna corresponde ao nome usado no pacote nlme.

Tabela 3.3. Exemplos de funções de variância que dependem de μ , mas não de $\boldsymbol{\delta}$ ^a

Função $g(\cdot)$	g_{ij}	Descrição
$\text{varPower}(\mu_{ij}, s_{ij}, \boldsymbol{\delta})$	$ \mu_{ij} ^{\delta_{s_{ij}}}$	potência de $ \mu_{ij} $, $\delta_{s_{ij}}$ conhecido
$\text{varExp}(\mu_{ij}, s_{ij}, \boldsymbol{\delta})$	$\exp(\mu_{ij}\delta_{s_{ij}})$	expoente de μ_{ij} , $\delta_{s_{ij}}$ conhecido
$\text{varConstPower}(\mu_{ij}, s_{ij}, \boldsymbol{\delta})$	$\delta_{1,s_{ij}} + \mu_{ij} ^{\delta_{2,s_{ij}}}$	constante mais função de variância potência, $\delta_{1,s_{ij}} > 0, \delta_{1,s_{ij}}$ e $\delta_{2,s_{ij}}$ conhecidos

^a Nome das funções usadas na primeira coluna corresponde ao nome usado no pacote nlme.

As funções de variância apresentadas na Tabela 3.1 dependem da variável re-

gressora da variância, isto é, v_{ij} , e dos parâmetros $\boldsymbol{\delta} = (\delta_1, \dots, \delta_S)$, mas não dependem de μ . Gałecki & Burzykowski (2013) referem-se a elas como funções de variância independentes da média. Observa-se também que a função $\text{varIdent}(\cdot)$ é definida somente para múltiplos estratos.

Na Tabela 3.2 são apresentados exemplos de funções de variância dependentes da média, pois dependem de μ_{ij} e de $\boldsymbol{\delta} = (\delta_1, \dots, \delta_S)$. Supondo uma forma particular para algumas das funções de variância apresentadas na Tabela 3.2, Gałecki & Burzykowski (2013) obtiveram funções de variância que dependem de μ , mas não de $\boldsymbol{\delta}$. Isto ocorre, por exemplo, para $\text{varPower}(\boldsymbol{\delta}, \mu_{ij}, s_{ij})$ ou $\text{varExp}(\boldsymbol{\delta}, \mu_{ij}, s_{ij})$ se for assumido $\delta_{s_{ij}} = 1$ (ou qualquer outra constante). Exemplos de funções de variância que dependem de μ , mas não dependem de $\boldsymbol{\delta}$ são apresentadas na Tabela 3.3.

Portanto, temos que as funções de variância permitem a modelagem de muitos padrões de heterocedasticidade. Por exemplo, usando $\text{varPower}(\mu_{ij}, s_{ij}, \boldsymbol{\delta})$ com $\delta_{s_{ij}} = 1$, obtemos $g(\mu_{ij}) = \mu_{ij}$ ou, equivalentemente, $\sigma_{ij} = \sigma\mu_{ij}$. Assim, obtemos um modelo no qual a variância mantém-se constante com os valores ajustados. Mas para $\delta_{s_{ij}} = 0,5$ obtemos um modelo no qual a variância aumenta linearmente com os valores ajustados.

3.3.2 Estrutura de correlação

Nesta seção apresentamos exemplos de estruturas para a matriz de correlação \mathbf{C}_i definida em (3.18). Pinheiro & Bates (2000) consideram que os erros ϵ_{ij} dentro da unidade amostral ou grupo são associados a vetores de posição \mathbf{p}_{ij} e que a correlação entre dois erros, ϵ_{ij} e $\epsilon_{ij'}$, depende dos correspondentes vetores de posição, \mathbf{p}_{ij} e $\mathbf{p}_{ij'}$, das variáveis regressoras somente através de alguma distância entre eles, digamos $d(\mathbf{p}_{ij}, \mathbf{p}_{ij'})$. Desta forma, a matriz de correlação \mathbf{C}_i fica especificada por

$$\text{Cor}(\epsilon_{ij}, \epsilon_{ij'}) = h[d(\mathbf{p}_{ij}, \mathbf{p}_{ij'}), \boldsymbol{\rho}], \quad i = 1, \dots, N, \quad j, j' = 1, \dots, n_i, \quad (3.21)$$

em que $\boldsymbol{\rho}$ é o vetor dos parâmetros de correlação, $d(\mathbf{p}_{ij}, \mathbf{p}_{ij'})$ é a função das distâncias dos vetores de posição \mathbf{p}_{ij} e $\mathbf{p}_{ij'}$, correspondentes a ϵ_{ij} e $\epsilon_{ij'}$, respectivamente, e $h(., .)$ é a função de correlação, contínua em relação a $\boldsymbol{\rho}$, assumindo valores entre -1 e 1 e $h(0, \boldsymbol{\rho}) = 1$.

Supondo várias distâncias e funções de correlação, é possível obter uma variedade de estruturas de correlação. Limitamos nossa discussão à classe de estruturas de correlação implementadas no `nlme`, cuja classificação é dada em estruturas seriais e estruturas espaciais.

Aqui, concentramos nossa atenção essencialmente às estruturas seriais que correspondem às estruturas de correlação definidas no contexto de dados longitudinais em relação ao tempo. Detalhes sobre estruturas espaciais, correspondentes às estruturas de correlação e definidas no contexto de dados espaciais, podem ser obtidos em Pinheiro & Bates (2000) ou em Gałecki & Burzykowski (2013), por exemplo.

Descrevemos na Tabela 3.4 as funções de correlação $h(., .)$, adaptadas de Gałecki & Burzykowski (2013), correspondentes às estruturas seriais. Os nomes das estruturas seguem a convenção usada no pacote `nlme`.

Tabela 3.4. Exemplos de estruturas de correlação serial

Estrutura de correlação ^a	Função $h(., .)$	Restrição
corCompSymm	$h(k, \boldsymbol{\rho}) = \rho$	$k = 1, 2, \dots; \rho < 1$
corAR1	$h(k, \boldsymbol{\rho}) = \rho^k$	$k = 0, 1, \dots; \rho < 1$
corCAR1	$h(s, \boldsymbol{\rho}) = \rho^s$	$s \geq 0; \rho \geq 0$
corSymm	$h(d(j, j'), \boldsymbol{\rho}) = \rho_{jj'}$	$j < j'; \rho_{jj'} < 1$

^a Nome das estruturas correspondem ao nome usado no pacote `nlme`.

Gałecki & Burzykowski (2013) chamam a atenção para um exemplo mais avançado de estrutura de correlação serial, a corARMA. Esta estrutura é obtida de um processo ARMA e descrita por $p + q$ parâmetros. Ao contrário das estruturas da Tabela 3.4, a função de correlação de um processo ARMA não possui forma explícita, sendo definida por uma relação recursiva. Para detalhes sobre a estrutura ARMA

veja Pinheiro & Bates (2000).

A estrutura de correlação serial mais simples, ou seja, simetria composta, referida como `corCompSymm`, assume mesma correlação entre todos os pares de erros dentro da unidade amostral, isto é

$$\text{Cor}(\epsilon_{ij}, \epsilon_{ij'}) = \rho, \quad (3.22)$$

que corresponde a (3.21) ao definir, $j \neq j'$ e $k = 1, 2, \dots$,

$$h(k, \boldsymbol{\rho}) = \rho. \quad (3.23)$$

Para as estruturas `corCompSymm` e `corAR1` a função distância é simplificada ainda mais se for assumido que a função depende da defasagem, ou seja, da diferença absoluta entre duas posições: $k = |p_{ij} - p_{ij'}|$. Para essas estruturas, a função $h(., .)$ depende simplesmente de k e ρ .

A estrutura de correlação geral ou não estruturada, `corSymm`, é o oposto da estrutura `corCompSymm`, no sentido de que geralmente leva a modelos superparametrizados. Com n_i observações em cada unidade amostral, a matriz de correlação não estruturada tem $n_i(n_i - 1)/2$ parâmetros. Se o número de parâmetros em $\boldsymbol{\theta}$ for relativamente grande, o processo de estimação $(\boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top)^\top$ pode se tornar instável, sobrando pouca informação para estimar $\boldsymbol{\beta}$, os parâmetros de maior interesse na modelagem. Portanto, esta estrutura só é viável quando N é grande em relação ao

$$\max_{1 \leq i \leq m} n_i.$$

3.3.3 Procedimento para seleção da estrutura de covariância \mathbf{R}_i

Uma etapa importante na modelagem da estrutura de covariância para matriz \mathbf{R}_i é como proceder para selecionar uma delas. Uma possível estratégia, utilizada em Littell et al. (2006), é avaliar a série de ajustes inspecionando critérios de informação, tais como o Critério de Informação de Akaike (*AIC*), (Akaike, 1974), em

que $AIC = -2LL + 2d$; Critério de Informação Bayesiano (BIC), (Schwarz, 1978), em que $BIC = -2LL + d \ln(n)$, e o AIC consistente, ($CAIC$), (Bozdogan, 1987), em que $CAIC = -2LL + d(1 + \ln(n))$, com LL representando o máximo da log-verossimilhança (completa ou restrita), d o número de parâmetros do modelo e n o número de observações. Outras referências interessantes para seleção de estrutura de covariância intra-unidades amostrais são Wolfinger (1993), Verbeke & Molenberghs (2000), Diggle et al. (2002), Rocha (2004) ou Rocha & Singer (2018), por exemplo.

Rocha (2004), com o objetivo de identificar uma estrutura de covariância para os dados, propôs um conjunto de técnicas gráficas e analíticas para o diagnóstico da matriz de covariâncias intra-unidades amostrais em estudos com medidas repetidas. Essas técnicas compreendem, entre outras, a análise da matriz de covariâncias e correlações amostrais, do gráfico de perfis individuais, do gráfico de perfis das linhas das matrizes de correlações amostrais em função da defasagem, do variograma amostral das observações padronizadas e, como uma ferramenta auxiliar, considera os critérios AIC e BIC .

No entanto, Rocha (2004) alerta que a utilização dessas técnicas, para a escolha dos modelos para a estrutura de covariância intra-unidades amostrais, depende de informações sobre o tipo de estudo com medidas repetidas (longitudinal ou não), sobre o planejamento (balanceado ou não) e sobre a existência de dados omissos, e propõe uma estratégia de seleção da estrutura de covariância intra-unidades amostrais para estudos não-longitudinais e outra para estudos longitudinais.

Singer et al. (2018) incorporam alguns itens à proposta sugerida por Rocha (2004) para a seleção da estrutura de covariância intra-unidades amostrais. Para estudos não-longitudinais a estratégia apresentada é:

1. Análise do planejamento do estudo para identificação de um modelo saturado para os efeitos fixos e de possíveis efeitos aleatórios.
2. Análise das matrizes de covariâncias e correlações amostrais, quando for possível calculá-las, para identificar padrões de variação.

3. Ajuste de modelos saturados para os parâmetros de localização identificados no item 1 com matrizes de covariâncias identificadas no item 2.
4. Utilização dos critérios *AIC* e *BIC* como ferramenta auxiliar na escolha da estrutura de covariância identificada nos itens 1 e 2.
5. Utilização de técnicas de diagnóstico para avaliar a qualidade do ajuste.
6. Ações corretivas para modificação do modelo.

E para estudos longitudinais é:

1. Análise do gráfico de perfis médios para propor um modelo para os efeitos fixos.
2. Análise dos gráficos de perfis individuais e de perfis individuais centralizados para identificar padrões que indiquem possíveis efeitos aleatórios e heterocedasticidade ao longo das condições de avaliação.
3. Análise das matrizes de covariâncias e correlações intra-unidades amostrais, quando for possível calculá-las, para identificar padrões de variação.
4. Análise da matriz de gráficos de dispersão das observações padronizadas quando for possível construí-los.
5. Análise de gráficos de linhas das matrizes de covariâncias e correlações amostrais em função das defasagens entre os instantes de observação.
6. Análise do variograma amostral como sugerido por Diggle et al. (2002).
7. No caso de ausência de dados omissos, análise do gráfico das linhas da matriz de covariâncias intra-unidades amostrais por meio de técnicas de regressão polinomial para a identificação de efeitos aleatórios.
8. No caso de presença de dados omissos, análise dos perfis individuais por meio de técnicas de regressão polinomial para a identificação de efeitos aleatórios.

9. Utilização dos critérios de informação como ferramenta auxiliar para a seleção das estruturas de covariância intra-unidades amostrais identificadas.
10. Utilização de técnicas de diagnóstico para avaliar a qualidade do ajuste.
11. Ações corretivas para modificação do modelo.

3.4 Testes para inclusão de efeitos aleatórios

Nos modelos mistos é de interesse testar se determinados efeitos aleatórios devem ser incluídos no modelo. Como sugere Drikvandi et al. (2013), de uma perspectiva estatística, testar a necessidade de inclusão efeitos aleatórios se traduz em testar a hipótese de que todos ou alguns dos componentes de variância da matriz \mathbf{D} são iguais a zero.

Ocorre que quando a hipótese nula está na fronteira do espaço paramétrico há uma violação nas condições de regularidade, como especificado em Cox & Hinkley (1974). Com isso não temos a garantia de que a distribuição assintótica da estatística da razão de verossimilhanças generalizada (RVG) seja a distribuição qui-quadrado, por exemplo.

No caso em que temos um único efeito aleatório no modelo, Self & Liang (1987) obtiveram a distribuição assintótica da estatística da RVG, sob a hipótese nula, para variáveis aleatórias independentes e identicamente distribuídas, dada pela mistura

$$0,50\chi_0^2 + 0,50\chi_1^2, \quad (3.24)$$

em que χ_0^2 denota a distribuição qui-quadrado com zero graus de liberdade (massa total no ponto zero) e χ_1^2 denota uma distribuição qui-quadrado com um grau de liberdade.

Stram & Lee (1994, 1995), usando os resultados de Self & Liang (1987), determinaram a distribuição assintótica da estatística da RVG para testes de hipóteses relativos à matriz \mathbf{D} , mostrando que na comparação de modelos encaixados, um com

q e outro com $q + 1$ efeitos aleatórios, a estatística da RVG deverá ser comparada a mistura

$$0,50\chi_q^2 + 0,50\chi_{(q+1)}^2, \quad (3.25)$$

em que χ_q^2 e $\chi_{(q+1)}^2$ denotam distribuições qui-quadrado com q e $q + 1$ graus de liberdade, respectivamente.

Para um modelo linear misto com apenas o intercepto fixo e um efeito aleatório, Crainiceanu & Ruppert (2004) mostraram que, fixado N e $n_i \rightarrow \infty$, a distribuição assintótica da estatística da RVG, sob a hipótese nula, é

$$(1 - a_N)\chi_0^2 + a_N\chi_1^2, \quad (3.26)$$

em que a_N é uma função de N .

Giampaoli & Singer (2009), a partir dos resultados de Vu & Zhou (1997) que generalizaram os resultados de Stram & Lee (1994, 1995) para situações em que as variáveis aleatórias não são independentes e identicamente distribuídas, concluíram que as distribuições assintóticas obtidas por Stram & Lee (1994, 1995) continuam válidas para modelos com até dois efeitos aleatórios e mostram a dificuldade em se verificar as condições de regularidade exigidas para além disso.

Nobre et al. (2008) reforçam que a principal desvantagem dos testes baseados nos métodos de verossimilhança generalizados (que são assintoticamente equivalentes aos testes do tipo score) consiste exatamente nessa dificuldade com relação às condições de regularidade exigidas e propuseram um teste para componentes de variância em modelos de efeitos aleatórios com um fator, baseado em decomposições de estatísticas U . Um aspecto importante no teste U proposto é que a derivação do teste não é afetada por tais dificuldades, além de poder servir como base para uma configuração mais geral, conforme indicado em Nobre (2007).

Lee & Braun (2012) discutem que embora tenha sido provado que a distribuição assintótica correta é uma mistura de distribuições χ^2 , a distribuição de mistura apropriada é bastante complexa e não intuitiva quando as hipóteses nula e alternativa

diferem em mais de um efeito aleatório. Como alternativa, Lee & Braun (2012) propuseram dois testes de permutação, um baseado nos BLUPs e outro baseado na estatística da razão de verossimilhanças restrita generalizada (RVRG). Ambos os testes não exigem dados balanceados, nem as medições precisam ocorrer nas mesmas condições de avaliação.

Essencialmente, os métodos envolvem resíduos ponderados com os pesos determinados pelas componentes de variância entre e dentro da unidade amostral. Lee & Braun (2012) calcularam a distribuição nula da permutação, para ambas as estatísticas, através da permutação dos resíduos e concluíram que são válidas assintoticamente e para amostras pequenas. Através de simulações, sob diferentes configurações, demonstraram que as perdas de poder de ambos os testes são pequenas e, em todas as configurações, exibem maior poder do que o teste assintótico.

3.4.1 Teste de permutação baseado nos BLUPs

O teste baseado nos BLUPs é utilizado unicamente para testar a inclusão de um efeito aleatório por vez e tem sua estatística de teste dada por

$$T_k = \sum_{i=1}^N \tilde{b}_{ik}^2 / N, \quad k = 1, \dots, q, \quad (3.27)$$

que é a variância amostral dos BLUPs para o efeito aleatório b_i . A estatística (3.27) decorre da soma dos BLUPs ao quadrado, que são tratados como uma amostra aleatória de $b_i \sim \mathcal{N}(0, \sigma_{b_i}^2)$.

Para exemplificar este método os autores consideraram testar as seguintes hipóteses:

$$H_0 : Y_{ij} = \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + \epsilon_{ij}, \quad (3.28)$$

$$H_1 : Y_{ij} = \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + \epsilon_{ij},$$

e

$$H_0 : Y_{ij} = \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + \epsilon_{ij}, \quad (3.29)$$

$$H_1 : Y_{ij} = \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + b_{i2} z_{2ij} + \epsilon_{ij}.$$

Para testar o conjunto de hipóteses (3.28) a estatística de teste é a expressão dada em (3.27) em que $k = 1$.

Para construir a distribuição das permutações para comparar com a estatística de teste observada, o processo se baseia nas permutações dos erros marginais, $\mathbf{e} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$. Sob a hipótese nula de que $\sigma_{b_{i1}}^2 = 0$, os erros são permutáveis e, mais especificamente, independentes e identicamente distribuídos, em que

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}). \quad (3.30)$$

Os erros têm o benefício de não exigir que os \mathbf{X}' s contínuos sejam idênticos entre todas as unidades amostrais, nem o número de observações para cada unidade amostral precisa ser o mesmo. A partir dessas afirmativas, os autores asseguram que podemos permutar os erros tanto entre quanto dentro das unidades amostrais. Na prática, os erros são preditos pelos resíduos $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, calculados a partir de estimativas dos parâmetros do modelo sob a hipótese alternativa de (3.28). Em acréscimo, os autores comentam que os resíduos também são assintoticamente permutáveis entre e dentro das unidades amostrais sob a hipótese nula.

Os resíduos marginais fazem parte do cálculo dos BLUPs e levam diretamente a uma distribuição de permutação para T_1 . Os autores utilizam $v = 1, 2, \dots, 1000$ permutações. Para cada permutação os resíduos marginais são permutados aleatoriamente. Usando esses resíduos permutados é gerada a estimativa $\hat{\sigma}_{b_i, v}^2$ referente a v -ésima permutação para $\sigma_{b_i}^2$, a partir da qual são calculados os valores permutados dos BLUPs que serão usados para a obtenção de T_{1v}^* , que é o valor permutado da estatística de teste T_1 . Com esses 1000 valores permutados de T_1 constrói-se uma distribuição nula empírica aproximada para T_1 . A re-estimação de $\sigma_{b_i}^2$ é feita considerando que algumas permutações dos resíduos resultarão em $\hat{\sigma}_{b_i}^2 = 0$ e levarão a uma distribuição nula empírica com massa positiva em zero. A probabilidade de significância é obtida calculando a proporção de T_1^* maiores que T_1 .

Quando o interesse é testar a presença de um único efeito aleatório em um mo-

delo que contém outros efeitos aleatórios, conforme a configuração (3.29), o modelo associado à hipótese nula contém um efeito aleatório tal que os $n = \sum_{i=1}^N n_i$ erros não são mais permutáveis sob essa hipótese. Em vez disso tem-se

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_0), \quad (3.31)$$

em que $\mathbf{V}_0 = \sigma_{b_{10}}^2 \mathbf{Z}\mathbf{Z}^\top + \boldsymbol{\Sigma}_0$, com $\boldsymbol{\Sigma}_0 = \sigma_{\epsilon_0}^2 \mathbf{I}$.

Esse problema é resolvido ponderando-se os erros pela matriz $(\mathbf{U}_0^\top)^{-1}$, em que \mathbf{U}_0 é a decomposição de \mathbf{V}_0 utilizando o método de Cholesky, isto é, $\mathbf{V}_0 = \mathbf{U}_0^\top \mathbf{U}_0$. Os autores afirmam que, como resultado, o conjunto de erros ponderados é

$$(\mathbf{U}_0^\top)^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (3.32)$$

sendo portanto permutáveis, permitindo novamente permutações tanto entre quanto dentro das unidades amostrais. Dessa forma, os autores expressam a estatística de teste (3.27), com $k = 2$, incorporando a decomposição de Cholesky, por

$$T_2 = \sum_{i=1}^N \tilde{b}_{i2}^2 / N = \sum_{i=1}^N \left[\boldsymbol{\Delta}_1 \mathbf{z} \hat{\mathbf{V}}_1^{-1} \hat{\mathbf{U}}_0^\top (\hat{\mathbf{U}}_0^\top)^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right]^2 / N. \quad (3.33)$$

Note que T_2 é calculada apenas para o único efeito aleatório cuja inclusão está sendo testada. Para os dados observados, a estatística de teste permanece sendo a variância amostral de \tilde{b}_{i2} , pois $\mathbf{U}_0^\top (\mathbf{U}_0^\top)^{-1}$ é igual à identidade para os resíduos ponderados não permutados. Os autores afirmam que com pesos apropriados o teste de permutação baseado nos BLUPs pode ser utilizado para realizar inferências quando se tem interesse em testar a inclusão de um único efeito aleatório.

Lee & Braun (2012) verificaram, através de estudos de simulação, que este teste de permutação é válido e tem poder comparável ao teste da RVG, cuja distribuição assintótica é uma mistura de qui-quadrados. O teste baseado nos BLUPs tem uma limitação: pode testar apenas a inclusão de um efeito aleatório de cada vez.

Na Subseção 3.4.2 está apresentado o teste de permutação baseado na estatís-

tica da RVRG que permite o teste para inclusão de múltiplos efeitos aleatórios e do qual o teste de permutação baseado nos BLUPs é um caso especial.

3.4.2 Teste de permutação baseado na RVRG

Esse teste de permutação é baseado na estatística da RVRG,

$$\lambda = -2 \ln(L_{H_0} - L_{H_1}), \quad (3.34)$$

em que L_{H_0} e L_{H_1} são as verossimilhanças sob as hipóteses nula e alternativa, respectivamente. Usando a mesma notação do modelo (3.8), em que $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$ e $\mathbf{e} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, tem-se

$$\lambda = \ln[|\mathbf{V}_0|/|\mathbf{V}_1|] + \mathbf{e}^\top (\mathbf{V}_0^{-1} - \mathbf{V}_1^{-1}) \mathbf{e} + \ln[|\mathbf{X}^\top \mathbf{V}_0^{-1} \mathbf{X}|/|\mathbf{X}^\top \mathbf{V}_1^{-1} \mathbf{X}|], \quad (3.35)$$

em que $|\mathbf{A}|$ denota o determinante da matriz \mathbf{A} .

Para as hipóteses em (3.28) os autores constroem a estatística da RVRG, que também envolve os erros marginais, sendo dada por

$$T_3 = \ln[|\widehat{\mathbf{V}}_0|/|\widehat{\mathbf{V}}_1|] + \widehat{\mathbf{e}}_1^\top (\mathbf{V}_0^{-1} - \mathbf{V}_1^{-1}) \widehat{\mathbf{e}}_1 + \ln[|\mathbf{X}^\top \widehat{\mathbf{V}}_0^{-1} \mathbf{X}|/|\mathbf{X}^\top \widehat{\mathbf{V}}_1^{-1} \mathbf{X}|], \quad (3.36)$$

com todos os parâmetros de λ (3.35) substituídos por suas estimativas sob as hipóteses nula e alternativa como denotado por seus índices.

Semelhantemente ao teste de permutação baseado nos BLUPs, novas matrizes $\widehat{\mathbf{V}}_0$ e $\widehat{\mathbf{V}}_1$ são estimadas para cada permutação de $\widehat{\mathbf{e}}_1$ e denotadas por $\widehat{\mathbf{V}}_0^*$ e $\widehat{\mathbf{V}}_1^*$, respectivamente. Os resíduos permutados são tratados como uma variável resposta e tanto $\widehat{\mathbf{V}}_0^*$ como $\widehat{\mathbf{V}}_1^*$ são estimadas a partir de um modelo misto com intercepto e inclinação aleatórios fixados na hipótese nula e a hipótese alternativa, respectivamente.

A reestimação de $\widehat{\mathbf{V}}_0^*$ e $\widehat{\mathbf{V}}_1^*$ é necessária devido às mudanças que ocorrem no posto de $\widehat{\boldsymbol{\Sigma}}$ quando as estimativas das variâncias dos efeitos aleatórios são iguais a zero. Lee & Braun (2012) asseguram que se \mathbf{V}_0 e \mathbf{V}_1 , incluindo $\boldsymbol{\Sigma}_0$ e $\boldsymbol{\Sigma}_1$, não

forem reestimadas, a distribuição da permutação será completamente baseada nas estimativas obtidas com os dados observados.

Ao testar a presença de um efeito aleatório com um ou mais efeitos aleatórios adicionais na hipótese nula, o procedimento é similar ao do teste de permutação baseado nos BLUPs. Para poder permutar os erros, eles antes devem ser ponderados por $(\mathbf{U}_0^\top)^{-1}$. Uma vez ponderados, os erros são permutáveis. Os erros ponderados permutáveis são então multiplicados por (\mathbf{U}_0^\top) para retornar à escala original dos resíduos e, para cada permutação, $\widehat{\mathbf{V}}_0^*$ e $\widehat{\mathbf{V}}_1^*$, são reestimados usando os erros permutáveis não ponderados, conforme descrito anteriormente. Então, T_3^* é calculada e a distribuição de permutação é gerada para a estatística da RVRG que será comparada com a estatística de teste observada, obtendo assim a probabilidade de significância.

Se a intenção for testar a inclusão de $0 < r \leq q$ efeitos aleatórios, o conjunto de hipóteses é

$$\begin{aligned} H_0 : Y_{ij} &= \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + \dots + b_{i(q-r)} z_{(q-r)ij} + \epsilon_{ij}, \\ H_1 : Y_{ij} &= \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + b_{i1} z_{1ij} + \dots + b_{iq} z_{qij} + \epsilon_{ij}. \end{aligned} \quad (3.37)$$

Os passos para a obtenção da estatística do teste para as hipóteses (3.37) são semelhantes aos das hipóteses (3.28) e (3.29). No entanto, Lee & Braun (2012) enfatizam a importância da reestimação de Σ_0 e Σ_1 após cada permutação, quando se testam a inclusão de múltiplos efeitos aleatórios, e afirmam que para um valor de r , a simulação é a única abordagem existente para calcular as probabilidades corretas de mistura para as distribuições qui-quadrado.

Embora ambos os métodos sejam computacionalmente intensivos, Lee & Braun (2012) argumentam que a computação paralela por meio da arquitetura para múltiplos núcleos (multi-core) torna possível reduzir o tempo necessário para implementação dos testes e finalizam destacando que a maior contribuição dos métodos consiste no fato de que o teste de permutação baseado na estatística da RVRG gera, automaticamente, as probabilidades corretas de mistura, já que o posto de $\widehat{\Sigma}^*$ muda de permutação para permutação.

3.5 Diagnóstico

Outro tópico importante no ajuste de um modelo de regressão é a análise de diagnóstico, pois possibilita a verificação de possíveis afastamentos das suposições feitas no modelo e a identificação de observações influentes ou extremas que podem vir a interferir nos resultados do ajuste. Esta seção reproduz os resultados de Singer et al. (2017).

Para o caso da avaliação de modelos lineares padrão têm-se muitas alternativas como a distância de Cook (1977), que originalmente foi desenvolvida para modelos lineares normais; o estudo da diagonal principal da matriz de projeção, apresentada por Hoaglin & Welsch (1978), que motivou a definição de pontos de alavanca; a definição dos vários tipos de padronização dos resíduos discutida por Belsley et al. (1980) e Cook & Weisberg (1982) ou os métodos de avaliação de influência conjunta das observações sob pequenas mudanças ou perturbações no modelo ou nos dados propostos por Cook (1986), por exemplo.

Para avaliar o ajuste de modelos lineares generalizados (MLG), as técnicas de diagnóstico também já se encontram muito bem definidas. Paula (2013) fez uma abrangente revisão dos métodos de diagnóstico em que define pontos de alavanca e discute análise de resíduos, métodos de deleção de pontos e influência local, dentre outros procedimentos.

Venezuela (2004) desenvolveu, baseando-se no trabalho de Tan et al. (1997), uma generalização das técnicas de diagnóstico para os MLG levando em consideração dados com medidas repetidas.

Singer et al. (2018) consideram que generalizações para os modelos lineares mistos são mais recentes citando, entre outros, Lesaffre & Verbeke (1998), Fung et al. (2002), Nobre & Singer (2007), Nobre & Singer (2011), Demidenko (2013) e Singer et al. (2017) que apresentam uma série de métodos de diagnóstico.

Com foco na análise de resíduos, mas também abordando as análises de influência global e influência local, Singer et al. (2017) descrevem e atualizam os métodos de

diagnósticos apresentados por Cook (1977), Belsley et al. (1980), Cook & Weisberg (1982), Cook (1986), Beckman et al. (1987), Chatterjee & Hadi (1988), Hilden-Minton (1995), Lesaffre & Verbeke (1998), Wei et al. (1998), Tan et al. (2001), Fung et al. (2002), Fei & Pan (2003), Demidenko & Stukel (2005), Nobre & Singer (2007), Zhu et al. (2007), Zewotir (2008), Nobre & Singer (2011), Schützenmeister & Pi-pho (2012) e Mun & Lindstrom (2013). E, com base na interpretação de gráficos de diagnósticos, também identificam medidas corretivas para possíveis violações às suposições assumidas ao modelo misto proposto.

Com base no modelo (3.8) é possível definir três tipos de erros,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \overbrace{\underbrace{\mathbf{Z}\mathbf{b}}_{\text{Efeito aleatório}} + \underbrace{\boldsymbol{\epsilon}}_{\text{Erro condicional}}}_{\text{Erro marginal (e)}},$$

mais especificamente,

- Erros marginais: $\mathbf{e} = \mathbf{Y} - \mathbb{E}(\mathbf{Y}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$,
- Erros condicionais: $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbb{E}(\mathbf{Y}|\mathbf{b}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$, e
- Efeitos aleatórios: $\mathbf{Z}\mathbf{b} = \mathbb{E}(\mathbf{Y}|\mathbf{b}) - \mathbb{E}(\mathbf{Y})$.

Os resíduos associados a cada tipo de erro, definidos em Hilden-Minton (1995), Verbeke & Lesaffre (1996) ou Pinheiro & Bates (2000), são dados por

- (i) Resíduos marginais: $\widehat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$,
- (ii) Resíduos condicionais: $\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{b}}$ e
- (iii) Resíduos de efeitos aleatórios: $\widehat{\mathbf{Z}\mathbf{b}} = \mathbb{E}(\mathbf{Y}|\widehat{\mathbf{b}}) - \mathbb{E}(\mathbf{Y})$,

sendo úteis para avaliar alguma suposição do modelo (3.8).

A seguir são apresentadas as ferramentas gráficas de diagnóstico construídas com base nos resíduos (i)–(iii). Acrescenta-se que os resultados e teoria aqui expostos, para avaliar a adequação do modelo ajustado, estão fortemente embasados nas ferramentas de diagnóstico propostas por Singer et al. (2017).

3.5.1 Resíduos marginais

Uma medida similar à proposta por Lesaffre & Verbeke (1998), para identificar unidades amostrais as quais a estrutura de covariância adotada, \mathbf{V}_i , parece não ser adequada, foi construída por Singer et al. (2017) e denominada de Índice de Lesaffre-Verbeke modificado. Tal medida tem a seguinte estrutura

$$\mathcal{V}_i^* = \frac{\sqrt{\mathcal{V}_i}}{n_i} = \frac{\sqrt{\|\mathbf{I}_{n_i} - \widehat{\mathbf{e}}_i^* \widehat{\mathbf{e}}_i^{*\top}\|^2}}{n_i}, \quad i = 1, \dots, N, \quad (3.38)$$

em que $\widehat{\mathbf{e}}_i^* = [\widehat{\mathbb{V}}(\widehat{\mathbf{e}}_i)]^{-1/2} \widehat{\mathbf{e}}_i$, são os resíduos marginais padronizados associados à i -ésima unidade amostral, com $\mathbb{V}(\widehat{\mathbf{e}}_i) = \mathbf{V}_i - \mathbf{X}_i (\mathbf{X}_i^\top \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top$. Vale salientar que a inclusão do denominador n_i em \mathcal{V}_i^* , pelos autores, tem por finalidade não imputar um peso demasiado às unidades amostrais com muitas observações.

Para valores de \mathcal{V}_i^* próximos de zero, o gráfico \mathcal{V}_i^* proposto em função do índice i indica para quais unidades a estrutura de covariância, \mathbf{V}_i , está adequada.

Singer et al. (2017) recomendam ainda a construção de mais dois gráficos considerando os componentes dos resíduos marginais padronizados dados pela expressão

$$\widehat{e}_{ij}^* = \frac{\widehat{e}_{ij}}{\left[\text{diag}_j \left(\widehat{\mathbb{V}}(\widehat{\mathbf{e}}_i) \right) \right]^{1/2}}, \quad (3.39)$$

na qual $\text{diag}_j(\mathbb{V}(\widehat{\mathbf{e}}_i))$ indica o j -ésimo elemento da diagonal principal de $\mathbb{V}(\widehat{\mathbf{e}}_i)$.

Gráficos de diagnóstico de \widehat{e}_{ij}^* em função dos valores de cada variável regressora, bem como em função dos valores ajustados, são sugeridos para avaliar a linearidade dos efeitos fixos, $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, e, de \widehat{e}_{ij}^* em função dos índices das observações, para detectar observações Y_{ij} discrepantes.

3.5.2 Resíduos condicionais

Segundo Hilden-Minton (1995) e posto por Nobre & Singer (2007), um resíduo é considerado puro, para um tipo de erro específico, se ele depende apenas das

componentes fixas e do erro que ele pretende predizer. Caso dependa de outros tipos de erros, é considerado um resíduo confundido. Sendo este o caso, por exemplo, dos resíduos $\widehat{\boldsymbol{\epsilon}}$ e $\mathbf{Z}\widehat{\mathbf{b}}$, uma vez que

$$\widehat{\boldsymbol{\epsilon}} = \boldsymbol{\Sigma}\mathbf{Q}\mathbf{Y} = \boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\epsilon} + \boldsymbol{\Sigma}\mathbf{Q}\mathbf{Z}\mathbf{b}, \quad (3.40)$$

$$\mathbf{Z}\widehat{\mathbf{b}} = \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^\top\mathbf{Q}\mathbf{Y} = \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^\top\mathbf{Q}\mathbf{Z}\mathbf{b} + \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^\top\mathbf{Q}\boldsymbol{\epsilon}, \quad (3.41)$$

sendo, portanto, confundidos com \mathbf{b} e $\boldsymbol{\epsilon}$, respectivamente. A exceção ocorrerá quando as colunas de \mathbf{Z} pertencerem ao espaço gerado pelas colunas de \mathbf{X} . Para detalhes e demonstrações destas propriedades, veja Nobre (2004).

Devido à presença deste confundimento, a avaliação da suposição de normalidade dos erros condicionais, utilizando os resíduos condicionais, pode ficar comprometida.

Para solucionar esta problemática, Hilden-Minton (1995) propõe a utilização de uma transformação linear nos resíduos condicionais a fim de minimizar a fração de confundimento, dada pela seguinte expressão

$$\mathbf{c}_k^\top\widehat{\boldsymbol{\epsilon}} = \lambda_k^{1/2}\boldsymbol{\ell}_k^\top\boldsymbol{\Sigma}^{-1/2}\mathbf{Y}, \quad k = 1, \dots, n - p$$

tal que $1 \geq \lambda_1 \geq \dots \geq \lambda_{n-p} \geq 0$ são os valores ordenados da matriz $\boldsymbol{\Lambda}$, obtidos da decomposição de valores singulares de

$$\boldsymbol{\Sigma}^{1/2}\mathbf{Q}\boldsymbol{\Sigma}^{1/2} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}^\top,$$

$\mathbf{c}_k = \lambda_k^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\ell}_k$ e $\boldsymbol{\ell}_k$ representa a k -ésima coluna de \mathbf{L} .

Nobre & Singer (2007) consideram a obtenção dos resíduos condicionais minimamente confundidos padronizados, especificamente por

$$\mathbf{c}_k^\top\widehat{\boldsymbol{\epsilon}}^* = \frac{\mathbf{c}_k^\top\widehat{\boldsymbol{\epsilon}}}{\mathbf{C}\widehat{\boldsymbol{\Sigma}}^\top\widehat{\mathbf{Q}}\widehat{\boldsymbol{\Sigma}}^\top\mathbf{C}^\top}, \quad (3.42)$$

e, por meio de estudos de simulação, concluíram que o gráfico quantil-quantil (QQ) gaussiano para $\mathbf{c}_k^\top \widehat{\boldsymbol{\epsilon}}^*$ pode ser utilizado para checar a adequação da distribuição normal proposta aos erros condicionais, mesmo quando os efeitos aleatórios não seguem uma distribuição normal.

Como forma de avaliar a homocedasticidade do erro condicional, Singer et al. (2017) sugerem construir um gráfico dos componentes dos resíduos condicionais padronizados, definidos na forma

$$\widehat{\boldsymbol{\epsilon}}_{ij}^* = \frac{\widehat{\boldsymbol{\epsilon}}_{ij}}{\text{diag}_{ij}(\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma})^{1/2}}, \quad (3.43)$$

com $\text{diag}_{ij}(\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma})$ denotando o elemento da diagonal principal de $\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma}$ que corresponde à j -ésima observação da i -ésima unidade amostral, em função dos valores ajustados. E, para detectar observações discrepantes, gráficos dos $\widehat{\boldsymbol{\epsilon}}_{ij}^*$ em função dos índices das observações.

3.5.3 Resíduos de efeitos aleatórios

Para identificar unidades amostrais discrepantes, Singer et al. (2017) sugerem a construção de gráficos baseados na distância de Mahalanobis, como proposto por Waternaux et al. (1989),

$$\mathcal{M}_i = \widehat{\mathbf{b}}_i^\top \left\{ \widehat{\mathbf{V}} \left[\widehat{\mathbf{b}}_i - \mathbf{b}_i \right] \right\}^{-1} \widehat{\mathbf{b}}_i, \quad (3.44)$$

em função dos índices das unidades amostrais.

A fim de avaliar a suposição de normalidade do vetor de efeitos aleatórios, \mathbf{b} , os autores propõem a construção do gráfico QQ baseado na distribuição χ_q^2 para \mathcal{M}_i .

Um resumo, similar ao que foi apresentado em Singer et al. (2017), apresentando os oito gráficos de diagnósticos discutidos acima encontra-se disponível na Tabela 3.5.

Tabela 3.5. Tipos de gráficos de diagnóstico baseados nos resíduos marginais, condicionais e de efeitos aleatórios, propostos em Singer et al. (2017), para avaliar o ajuste do modelo linear misto

Resíduo	Gráfico	Característica identificada
Marginal	\mathcal{V}_i^* vs índices das unidades	Unidades para as quais \mathbf{V}_i deveria ser modificada
	$\hat{\mathbf{e}}_{ij}^*$ vs variáveis regressoras ou valores ajustados	Linearidade dos efeitos fixos
	$\hat{\mathbf{e}}_{ij}^*$ vs índices das observações	Observações discrepantes
Condicional	QQ gaussiano para $\mathbf{c}_k^\top \hat{\boldsymbol{\epsilon}}^*$	Normalidade dos erros condicionais
	$\hat{\boldsymbol{\epsilon}}_{ij}^*$ vs valores ajustados	Homocedasticidade dos erros condicionais
	$\hat{\boldsymbol{\epsilon}}_{ij}^*$ vs índices das observações	Observações discrepantes
Efeitos aleatórios	QQ da distribuição χ_q^2 para \mathcal{M}_i	Normalidade dos efeitos aleatórios
	\mathcal{M}_i vs índices das unidades	Unidades discrepantes

Como no ajuste da regressão linear clássica, nos modelos lineares mistos os casos que apontam valores altos de resíduos (padronizados) não são necessariamente problemáticos para a inferência. Todavia, é importante verificar se o modelo é influenciado por uma ou um conjunto de observações, sobre as quais algumas das análises e conclusões podem sofrer considerável variação. Neste sentido, o modelo é sensível a essas observações e o procedimento estatístico claramente não é robusto (Fung et al., 2002; Nobre, 2004). Vale salientar ainda que a análise de influência é complementar à análise de resíduos.

3.5.4 Análise de pontos alavanca

O poder de alavanca (*leverage*) é um dos principais componentes na análise de influência em modelos de regressão linear, sendo medido pelos elementos h_{ii} da matriz de projeção $\mathbf{H} = \mathbf{X}(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}$, que expressam quão extremas são as observações no espaço das variáveis regressoras (Cook & Weisberg, 1982; Chatterjee & Hadi, 1988). Um poder de alavanca alto (*high leverage*) significa um maior potencial para

influenciar no ajuste do modelo.

Mediante devidas adaptações, a ideia do poder de alavanca foi generalizada, por exemplo, por Wei et al. (1998) que investigaram o poder de alavanca para modelos não lineares da família exponencial, que incluem a regressão não linear e os modelos lineares generalizados como seus casos especiais; Demidenko & Stukel (2005) que estenderam várias técnicas de diagnóstico comumente usadas em regressão linear para o modelo linear misto, incluindo o poder de alavanca; e Nobre & Singer (2011) que propuseram incorporar a informação dos valores condicionais ajustados e definiram a matriz de alavanca conjunta generalizada para modelos lineares mistos.

Com o objetivo de introduzir as informações dos efeitos aleatórios, Nobre & Singer (2011) consideraram $\widehat{\mathbf{Y}}^* = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\mathbf{b}}$. Assim, derivando $\widehat{\mathbf{Y}}^*$ com relação a \mathbf{Y}^\top , a matriz de alavanca conjunta generalizada fica dada por

$$\begin{aligned} \mathbf{L} &= \frac{\partial \widehat{\mathbf{Y}}^*}{\partial \mathbf{Y}^\top} = \frac{\partial \mathbf{X}\widehat{\boldsymbol{\beta}}}{\partial \mathbf{Y}^\top} + \frac{\partial \mathbf{Z}\widehat{\mathbf{b}}}{\partial \mathbf{Y}^\top} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} + \mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}^\top \mathbf{Q} \\ &= \mathbf{L}_1 + \mathbf{L}_2 \mathbf{Q}, \end{aligned} \quad (3.45)$$

em que \mathbf{L}_1 é a matriz de alavanca marginal generalizada, como sugerido por Fung et al. (2002), e a matriz \mathbf{L}_2 , seguindo Nobre & Singer (2011), representa a porção da variabilidade intra-unidades amostrais explicada pela presença dos efeitos aleatórios, denominada de matriz de alavanca de componentes aleatórios generalizada. A particularidade da decomposição da matriz (3.45) é a conveniência de examinar a influência conjunta de cada observação ou unidade tanto para os valores marginais como para os valores condicionais ajustados.

O i -ésimo bloco da diagonal principal de \mathbf{L}_1 é denotado por $\mathbf{L}_{1i} = \mathbf{X}_i(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}_i^\top \mathbf{V}_i^{-1}$. Nobre & Singer (2011) definem o j -ésimo elemento da diagonal principal de \mathbf{L}_{1i} e denotam por $\mathbf{L}_{1i(jj)}$. Desta forma,

$$\text{tr}(\mathbf{L}_1) = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbf{L}_{1i(jj)} = p. \quad (3.46)$$

Baseados no resultado (3.46), alguns critérios descritivos são considerados na literatura para dizer que a observação j ou a unidade amostral i tem alto poder de alavanca com relação aos valores marginais ajustados, $\hat{\mathbf{Y}}$. A medida $2p/n$ é um exemplo de um limite alternativo para comparar os elementos diagonais de \mathbf{L}_{1i} . Desta forma, observações com valores $\mathbf{L}_{1i(jj)} \geq 2p/n$, assim como unidades amostrais com valores $\text{tr}(\mathbf{L}_{1i})/n_i \geq 2p/n$, indicam alto poder de alavanca com relação aos valores marginais ajustados.

Nobre & Singer (2011) recomendam que a alternativa mais apropriada para avaliar o poder de alavanca das observações e unidades amostrais, com relação aos efeitos aleatórios associados aos valores condicionais ajustados, é utilizando a matriz de alavanca de componentes aleatórios generalizada \mathbf{L}_2 e apontam que a matriz proposta por Demidenko & Stukel (2005), para identificar observações com alto poder de alavanca para os componentes aleatórios dos valores condicionais ajustados, por ela depender da matriz \mathbf{L}_1 , pode mascarar o poder de alavanca a ser investigado.

Portanto, $\mathbf{L}_{2i} = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$ é definida como a matriz de alavanca de componentes aleatórios generalizada associada à i -ésima unidade e $\mathbf{L}_{2i(jj)}$ corresponde a matriz de alavanca de componentes aleatórios generalizada para a j -ésima observação da i -ésima unidade e denota o j -ésimo elemento da diagonal principal de \mathbf{L}_{2i} .

Para dizer quais unidades amostrais ou observações devem ser consideradas como tendo alto poder de alavanca com respeito aos componentes aleatórios dos valores condicionais ajustados, Nobre & Singer (2011) dão como exemplo de ponto de corte a medida $2\text{tr}(\mathbf{L}_2)/n$. Assim, convém examinar aquelas unidades amostrais com valores $\text{tr}(\mathbf{L}_{2i})/n_i \geq 2\text{tr}(\mathbf{L}_2)/n$ e observações com valores $\mathbf{L}_{2i(jj)} \geq 2\text{tr}(\mathbf{L}_2)/n$.

Dois pontos a destacar: primeiro, a definição, como destacado em Nobre (2004), da utilização das estimavas dos parâmetros fixos (BLUE) e as predições dos efeitos aleatórios (BLUP) para a computação dos valores da matriz \mathbf{L} ; segundo, diz respeito não haver um ponto de corte, ou limite, padrão a ser utilizado como critério para investigação de possíveis unidades amostrais ou observações com alto poder de alavanca, sendo, portanto, bastante razoável utilizar critérios subjetivos baseados na

inspeção visual.

3.5.5 Análise de omissão de casos

A omissão de casos (*case deletion*) sugere avaliar o impacto sobre as estimativas da regressão quando se retira uma ou um conjunto de observações do conjunto de dados. Uma observação é influente se o efeito de excluí-la do conjunto de dados causar diferenças substanciais na análise. Nesse sentido, a distância de Cook, introduzida por Cook (1977), é a técnica mais conhecida e utilizada.

No contexto dos modelos lineares mistos, como indicado em Singer et al. (2017), o conjunto de dados a ser omitido será o das medidas repetidas realizadas na mesma unidade amostral, denominada de **caso**, ou mesmo um conjunto de unidades amostrais. Nessas circunstâncias, a atualização do BLUE e BLUP, sem a necessidade de reajustar o modelo, pode ser trabalhosa e merece atenção.

Nobre (2004), seguindo Hilden-Minton (1995), apresenta a formulação para atualização do BLUE e BLUP quando se omite um conjunto de unidades amostrais $I = i_1, \dots, i_k$, em que $1 \leq i_1 \leq \dots \leq i_k \leq n$, e supondo $\mathbf{\Delta}$ e $\mathbf{\Sigma}$ conhecidas. Os autores mostram que os estimadores $\widehat{\boldsymbol{\beta}}_{(I)}$ e $\widehat{\mathbf{b}}_{(I)}$, respectivamente, o BLUE e o BLUP quando se omitem as observações do conjunto I , são similares ao BLUE de $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, e o BLUP de \mathbf{b} , $\widehat{\mathbf{b}}$, do modelo (3.8).

Resultados referentes aos estimadores $\widehat{\boldsymbol{\beta}}_{(I)}$ e $\widehat{\mathbf{b}}_{(I)}$ permitem considerar a seguinte generalização da distância de Cook

$$D_I = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)})^\top (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}) (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)})}{p} = \frac{(\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_{(I)})^\top \mathbf{V}^{-1} (\widehat{\mathbf{Y}} - \widehat{\mathbf{Y}}_{(I)})}{p} \quad (3.47)$$

como uma medida da influência das unidades do conjunto I na estimativa dos parâmetros fixos $\boldsymbol{\beta}$.

A abordagem da distância de Cook condicional também foi proposta como uma medida para a avaliação da influência (Tan et al., 2001). No entanto, considera um cenário em que o interesse recai no estudo do impacto da j -ésima observação da

i -ésima unidade amostral. Assim, a ocorrência de unidades amostrais identificadas como influentes pode ser devido ao efeito de uma única observação intra-unidade.

Considerando o modelo condicional, a distância de Cook condicional é definida como

$$\begin{aligned} D_{i(j)}^{cond} &= \sum_{i=1}^N \frac{\left(\widehat{\mathbf{Y}}_i^* - \widehat{\mathbf{Y}}_{i(j)}^*\right)^\top \mathbb{V}[\mathbf{Y}_i | \mathbf{b}_i]^{-1} \left(\widehat{\mathbf{Y}}_i^* - \widehat{\mathbf{Y}}_{i(j)}^*\right)}{(N-1)q + p} \\ &= \sum_{i=1}^N \frac{\left(\widehat{\mathbf{Y}}_i^* - \widehat{\mathbf{Y}}_{i(j)}^*\right)^\top \mathbf{R}_i^{-1} \left(\widehat{\mathbf{Y}}_i^* - \widehat{\mathbf{Y}}_{i(j)}^*\right)}{(N-1)q + p} \end{aligned} \quad (3.48)$$

em que $\widehat{\mathbf{Y}}_i^* = \mathbf{X}_i \widehat{\boldsymbol{\beta}} + \mathbf{Z}_i \widehat{\mathbf{b}}$, $\widehat{\mathbf{Y}}_{i(j)}^* = \mathbf{X}_i \widehat{\boldsymbol{\beta}}_{(i(j))} + \mathbf{Z}_i \widehat{\mathbf{b}}_{(i(j))}$ e $\widehat{\boldsymbol{\beta}}_{(i(j))}$ e $\widehat{\mathbf{b}}_{(i(j))}$ denotam, respectivamente, os BLUEs de $\boldsymbol{\beta}$ e \mathbf{b} obtidos após a omissão da j -ésima observação da i -ésima unidade amostral.

Para avaliar, separadamente, a influência sobre os parâmetros de efeitos aleatórios e de efeitos fixos, Tan et al. (2001) sugerem a decomposição de (3.48) em $D_{i(j)}^{cond} = D_{1i(j)}^{cond} + D_{2i(j)}^{cond} + D_{3i(j)}^{cond}$ e consideram esta uma grande vantagem. O primeiro termo resulta em

$$D_{1i(j)}^{cond} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))})^\top (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X}) (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))})}{(N-1)q + p} \quad (3.49)$$

sendo esta medida útil para a avaliar a influência da j -ésima observação da i -ésima unidade amostral na estimação de $\boldsymbol{\beta}$. O segundo termo em

$$D_{2i(j)}^{cond} = \frac{(\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))})^\top (\mathbf{Z}^\top \widehat{\mathbf{R}}^{-1} \mathbf{Z}) (\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))})}{(N-1)q + p} \quad (3.50)$$

sendo apropriado para a avaliar a influência da j -ésima observação da i -ésima unidade amostral na estimação de \mathbf{b} e o terceiro termo

$$D_{3i(j)}^{cond} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i(j))})^\top (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z}) (\widehat{\mathbf{b}} - \widehat{\mathbf{b}}_{(i(j))})}{(N-1)q + p} \quad (3.51)$$

está relacionado com mudanças na covariância entre as estimativas de β e \mathbf{b} quando a j -ésima observação da i -ésima unidade amostral é removida da análise.

Apresentamos na Tabela 3.6 um resumo para construção de gráficos de influência, adaptado de Singer et al. (2018), baseados nas medidas discutidas nas Subseções 3.5.4 e 3.5.5.

Para efeito do estudo dos gráficos de diagnóstico, trabalhamos com os limites de referência utilizados em Singer et al. (2017), que são o terceiro quartil mais 1,5 vezes o intervalo interquartilico. Nos gráficos de pontos alavanca, também pode-se incluir uma linha correspondente a $2p/N$.

Para gerar os gráficos elencados nas Tabelas 3.5 e 3.6 pode-se utilizar as funções desenvolvidas na linguagem R por Singer et al. (2017), disponíveis em <http://www.ime.usp.br/~jmsinger/lmmdiagnositics.zip>.

Tabela 3.6. Tipos de gráficos de influência baseados na decomposição da distância de Cook condicional e na alavanca generalizada no contexto do modelo linear misto

Medida de influência	Gráfico de índices	Avaliar a influência
Alavanca marginal generalizada \mathbf{L}_1	$\frac{\text{tr}(\mathbf{L}_{1i})}{n_i}$ vs unidades	Componente fixo de valor ajustado ($\mathbf{X}\hat{\beta}$)
	$\mathbf{L}_{1i(jj)}$ vs observações	
Alavanca para componentes aleatórios generalizada \mathbf{L}_2	$\frac{\text{tr}(\mathbf{L}_{2i})}{n_i}$ vs das unidades	Componente aleatório de valor ajustado ($\mathbf{Z}\hat{\mathbf{b}}$)
	$\mathbf{L}_{2i(jj)}$ vs das observações	
Distância de Cook condicional	$D_{1i(j)}^{cond}$ vs observações	Coefficientes de regressão ($\hat{\beta}$)
	$D_{2i(j)}^{cond}$ vs observações	Efeitos aleatórios ($\hat{\mathbf{b}}$)
	$D_{3i(j)}^{cond}$ vs observações	Mudanças na covariância entre $\hat{\beta}$ e $\hat{\mathbf{b}}$

3.6 Outras distribuições para os efeitos aleatórios

Embora os modelos lineares mistos sejam considerados uma ferramenta poderosa na modelagem estatística, tendo uma extensa teoria desenvolvida, tais desenvolvimentos, em grande parte, são baseados na suposição de normalidade para os

efeitos aleatórios e para os erros. No entanto, devido o aparecimento de fenômenos cada vez mais complexos e a falta de robustez da distribuição normal para as fontes de variação, o pressuposto de normalidade pode não ser adequado.

Uma possibilidade é a flexibilização do pressuposto de normalidade. Neste sentido, extensões do modelo (3.1)—(3.2) com distribuições alternativas têm sido consideradas, tais como a classe das distribuições elípticas que podem apresentar caudas mais leves ou mais pesadas que as caudas da distribuição normal. Pinheiro et al. (2001) propuseram um modelo linear hierárquico de efeitos mistos robusto assumindo que os efeitos aleatórios e os erros intra-unidades seguem distribuições t multivariadas; Savalli et al. (2006) introduziram os modelos lineares mistos elípticos com ênfase na avaliação dos componentes de variância; e Osorio et al. (2007) consideraram, em detalhes, análises de influência local sob vários esquemas de perturbação para modelos lineares elípticos com estrutura longitudinal.

Ibacache-Pulgar et al. (2012) introduziram a classe dos modelos mistos semi-paramétricos elípticos. Esses modelos, além de terem a flexibilidade de estender a classe dos erros para as distribuições elípticas, constituem uma alternativa para explicar a dependência das medidas intra-unidades amostrais através dos efeitos das variáveis explicativas que podem contribuir tanto de maneira paramétrica como não paramétrica.

Embora a classe das distribuições elípticas representem uma boa alternativa ao modelo normal, ela não é adequada em situações nas quais a distribuição das observações é assimétrica. Arellano-Valle et al. (2005), considerando que os efeitos aleatórios e os erros seguem uma distribuição normal assimétrica, propuseram o modelo linear misto normal assimétrico (SNLMM — *Skew-Normal Linear Mixed Model*). Lachos et al. (2010) apresentaram uma nova classe de modelos lineares mistos assimétricos em que os efeitos aleatórios seguem uma distribuição normal assimétrica multivariada e os erros seguem uma distribuição normal simétrica, por exemplo.

4 POLINÔMIOS FRACIONÁRIOS NOS MODELOS MISTOS

4.1 Introdução

Diversas estratégias podem ser empregadas para o ajuste do modelo PFMM, mas algumas podem vir a falhar devido a problemas de convergência dos métodos iterativos requeridos. Um polinômio no qual a própria potência é um parâmetro, caracteriza um modelo não linear cujo ajuste pode ser um desafio também devido a problemas de convergência, conforme discutido por autores clássicos no tema como Bates & Watts (1988) e Ratkowsky (1983), por exemplo.

Na presença de efeitos aleatórios, o modelo é não linear misto com desafios ainda maiores e, possivelmente, não aplicável na prática quando o número de variáveis regressoras excede três ou quatro, digamos. Mesmo explorando estratégias similares ao MFP, que usa o ferramental de ajuste de preditores lineares, ainda podemos nos deparar com desafios a resolver. Como observado em Singer et al. (2017), inclusive para o ajuste de modelos polinomiais convencionais para dados longitudinais, podemos observar problemas de convergência causados pelo possível mau condicionamento das matrizes de especificação, X e Z , do modelo.

Considerando estudos observacionais, que em geral contém diversas variáveis regressoras, propomos uma estratégia de ajuste para selecionar o modelo PFMM. Seguindo as recomendações de autores consagrados, nossa proposta se inicia com o modelo que envolve o número máximo de efeitos fixos possíveis de serem considerados (Verbeke & Molenberghs, 2000) e, seguindo o princípio da hierarquia (Wu & Hamada, 2009) a estratégia proposta compreende as seguintes etapas:

- (1) Seleção de uma estrutura média preliminar.
- (2) Seleção dos efeitos aleatórios.
- (3) Seleção de uma estrutura de covariância para os erros.
- (4) Redução do modelo.

A estratégia proposta é investigada, no Capítulo 5, na modelagem de três conjuntos de dados, cada um com determinada configuração no que diz respeito ao número e tipo de variáveis regressoras:

Aplicação 1: Uma variável regressora contínua.

Aplicação 2: Uma variável regressora contínua e uma binária.

Aplicação 3: Doze variáveis regressoras contínuas e uma binária.

Para a Aplicação 1, a mais simples, usamos dados de um estudo para modelar tendências não lineares em dados longitudinais que permite explicitar alguns pontos sutis da análise FP univariada que não foram abordados em outros trabalhos. A Aplicação 2, também com dados longitudinais, permite-nos estender o procedimento MFPI no contexto de modelos mistos, haja vista a investigação de interação entre as duas variáveis regressoras. A Aplicação 3, utilizando dados com uma estrutura hierárquica, possibilita-nos adaptar o procedimento MFP para a seleção de variáveis e determinação da forma funcional de cada regressora selecionada. Em todas as aplicações investigamos também a necessidade de modelagem da estrutura de variâncias e covariâncias.

Na Seção 4.2 apresentamos a construção do modelo PFMM e na Seção 4.3 descrevemos as etapas de (1) a (4) da estratégia proposta para o ajuste do modelo. Similarmente a Royston e coautores, nesta tese focaremos apenas nas famílias FP1 e FP2.

4.2 O Modelo PFMM

Sejam y_{ij} , x_{uij} e w_{vij} , os valores da variável resposta, u -ésima variável regressora contínua e v -ésima variável regressora binária ou *dummy*, respectivamente, medidas na j -ésima condição de avaliação da i -ésima unidade amostral, em que $i = 1, \dots, N$, $u = 1, \dots, U$, $v = 1, \dots, V$ e $j = 1, \dots, n_i$, com $n = \sum_{i=1}^N n_i$ sendo o total de observações e $L = U + V$ o total de variáveis regressoras. Acomodando o preditor de (2.5) em um modelo misto incorporando os possíveis componentes de efeitos aleatórios dos termos das variáveis regressoras, teremos o modelo PFMM. O modelo PFMM de grau 1 (PFMM1) proposto é expresso por

$$\begin{aligned}
 y_{ij} &= \beta_0 + \beta_1 x_{1ij}^{p^1} + \beta_2 x_{2ij}^{p^2} + \dots + \beta_U x_{Uij}^{p^U} + \gamma_1 w_{1ij} \\
 &\quad + \gamma_2 w_{2ij} + \dots + \gamma_V w_{Vij} + b_{i0} + b_{i1} x_{1ij}^{p^1} \\
 &\quad + b_{i2} x_{2ij}^{p^2} + \dots + b_{iU} x_{Uij}^{p^U} + \epsilon_{ij} \\
 &= \underbrace{\beta_0 + \sum_{u=1}^U \beta_u x_{uij}^{p^u} + \sum_{v=1}^V \gamma_v w_{vij}}_{\text{efeitos fixos}} + \underbrace{b_{i0} + \sum_{u=1}^U b_{iu} x_{uij}^{p^u}}_{\text{efeitos aleatórios}} + \epsilon_{ij},
 \end{aligned} \tag{4.1}$$

em que p^u é a potência da u -ésima variável regressora contínua e pertence ao conjunto $S = \{-2; -1; -0,5; 0; 0,5; 1; 2; 3\}$, com $p^u = 0$ indicando o logaritmo natural de x_{uij} , isto é, $x_{uij}^0 = \ln x_{uij}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_U, \gamma_1, \dots, \gamma_V)^\top$ é o vetor $((L + 1) \times 1)$ de parâmetros fixos, $\mathbf{b}_i = (b_{i0}, b_{i1}, \dots, b_{iU})^\top$ é o vetor $((U + 1) \times 1)$ de efeitos aleatórios e ϵ_{ij} representando o erro aleatório.

Seja uma transformação FP2 em x_{uij} com potências $\mathbf{p}^u = (p_1^u, p_2^u)$ ou (p_1^u, p_1^u) , conforme definido em (2.3), e seja $\boldsymbol{\beta}_u = (\beta_{u1}, \beta_{u2})^\top$, $\mathbf{b}_{iu} = (b_{iu1}, b_{iu2})^\top$ com $u =$

$1, \dots, U$. O modelo PFMM de grau 2 (PFMM2) proposto é expresso por

$$\begin{aligned}
y_{ij} &= \beta_0 + \beta_{11}x_{1ij}^{p_1^1} + \beta_{12}x_{1ij}^{p_2^1} + \beta_{21}x_{2ij}^{p_1^2} + \beta_{22}x_{2ij}^{p_2^2} + \dots + \beta_{U1}x_{Uij}^{p_1^U} + \beta_{U2}x_{Uij}^{p_2^U} \\
&\quad + \gamma_1w_{1ij} + \gamma_2w_{2ij} + \dots + \gamma_Vw_{Vij} + b_{i0} + b_{i11}x_{1ij}^{p_1^1} + b_{i12}x_{1ij}^{p_2^1} \\
&\quad + b_{i21}x_{2ij}^{p_1^2} + b_{i22}x_{2ij}^{p_2^2} + \dots + b_{iU1}x_{Uij}^{p_1^U} + b_{iU2}x_{Uij}^{p_2^U} + \epsilon_{ij} \\
&= \beta_0 + \underbrace{\sum_{u=1}^U x_{uij}^{\mathbf{p}^u} \boldsymbol{\beta}_u}_{\text{efeitos fixos}} + \underbrace{\sum_{v=1}^V \gamma_v w_{vij} + b_{i0} + \sum_{u=1}^U x_{uij}^{\mathbf{p}^u} \mathbf{b}_{iu}}_{\text{efeitos aleatórios}} + \epsilon_{ij}.
\end{aligned} \tag{4.2}$$

Note que ambas expressões, (4.1) e (4.2), apenas representam duas possíveis especificações do modelo PFMM2. Dependendo da aplicação, o modelo pode ser muito mais geral, incluindo, por exemplo, termos de interações principalmente na parte fixa, e outros termos na parte aleatória, inclusive com variáveis regressoras que não aparecem na parte fixa. Prezando por simplicidade de notação, a especificação geral não será explicitada.

Uma vez que uma função FP1 ou FP2 específica é incorporada ao modelo misto, os parâmetros do modelo PFMM são estimados usando os métodos descritos no Capítulo 3, podendo ser ajustado por máxima verossimilhança e máxima verossimilhança restrita.

Para a seleção do modelo PFMM propomos uma estratégia de ajustamento buscando uma modelagem apropriada para a estrutura média, bem como uma especificação adequada para a estrutura de covariância dos dados. Dentro desse processo, as ferramentas gráficas para possível detecção de afastamento das suposições do modelo, conforme desenvolvimento de Singer et al. (2017), podem ser utilizadas.

A estratégia considerada nessa tese segue as orientações gerais para construção de modelos mistos sugeridas por Verbeke & Molenberghs (2000), sendo inseridas adequações para incorporação dos procedimentos específicos dos FPs. As etapas da estratégia são descritas na Seção 4.3.

4.3 Estratégia de ajuste do modelo PFMM

Tanto a estrutura de média como a de covariância influenciam o ajuste geral, de modo que normalmente uma é mantida constante enquanto se investiga a outra (Verbeke & Molenberghs, 2000). A estratégia proposta, em cada etapa, segue essa orientação. Dependendo da complexidade do problema, ou seja, uma única variável regressora, múltiplas variáveis regressoras, com ou sem interação, um algoritmo FP específico é utilizado.

4.3.1 Etapa 1: Seleção de uma estrutura média preliminar

O primeiro passo da análise consiste em selecionar um modelo para a média da resposta. Devemos então iniciar com uma estrutura fixa bem especificada, isto é, a estrutura mais completa possível, a qual nos remete ao modelo saturado, no caso de este ser de ajuste viável. No entanto, como nosso interesse está nas variáveis regressoras contínuas, que acreditamos afetar a resposta média, o conceito de um modelo saturado desmorona, principalmente no caso de dados observacionais e a escolha de uma estrutura fixa bem especificada torna-se menos óbvia.

Nossa proposta para o processo de busca por uma estrutura média bem especificada para os efeitos fixos do modelo PFMM é iniciar apenas com o intercepto aleatório e sem estrutura de dependência nos erros ao avaliar o impacto dos vários FPs no ajuste global. Portanto, o modelo PFMM1 fica expresso por

$$y_{ij} = \beta_0 + \sum_{u=1}^U \beta_u x_{uij}^{p^u} + \sum_{v=1}^V \gamma_v w_{vij} + b_{i0} + \epsilon_{ij}, \quad (4.3)$$

e o PFMM2 por

$$y_{ij} = \beta_0 + \sum_{u=1}^U x_{uij}^{p^u} \beta_u + \sum_{v=1}^V \gamma_v w_{vij} + b_{i0} + \epsilon_{ij}, \quad (4.4)$$

em que $b_{0i} \sim N(0, \sigma_b^2)$ e $\epsilon_{ij} \sim N(0, \sigma^2)$. Dado o conjunto de potências S tem-se $8 \times U$

modelos possíveis para (4.3) e $36 \times U$ modelos possíveis para (4.4). A estimação dos parâmetros para cada um dos modelos ajustados é realizada pelo método de máxima verossimilhança. Devemos observar, no entanto, que dependendo da configuração das variáveis regressoras que compõem o modelo, a seleção do melhor modelo PFMM seguirá procedimentos diferentes.

Utilizaremos a notação $\text{PFMM1}(p)$ e $\text{PFMM2}(p_1, p_2)$ para apresentar as potências selecionadas dos respectivos modelos. Para a Aplicação 1, com apenas uma variável regressora contínua, os modelos $\text{PFMM1}(p)$ e $\text{PFMM2}(p_1, p_2)$ com melhor ajuste serão aqueles para o qual o vetor \mathbf{p} resultar na maior verossimilhança. Para seleção entre a melhor função $\text{PFMM1}(p)$ e $\text{PFMM2}(p_1, p_2)$ o procedimento de seleção da função (FSP), discutido na Seção 2.2.2, foi estendido para o modelo misto.

No caso da Aplicação 2, com uma variável regressora contínua e uma categórica, adaptamos o procedimento MFPI, descrito na Seção 2.4, para a investigação de possível interação entre as variáveis regressoras, lembrando que o passo 1 do algoritmo MFPI inicia com o melhor modelo $\text{PFMM2}(p_1, p_2)$ selecionado previamente.

E, no caso da aplicação 3, com diversas variáveis regressoras contínuas, são realizadas, obedecendo os passos do algoritmo MFP (Seção 2.3.1) adaptado aos modelos mistos, as seleções de variáveis regressoras significativas e a determinação da forma funcional. O processo de eliminação *backward*, BE, é combinado com o algoritmo FSP que seleciona a melhor transformação FP para cada variável regressora contínua, por vez. Dependendo do valor-p associado à melhor transformação, uma ou mais variáveis regressoras podem ser excluídas do modelo final (Royston et al., 1999).

Para implementação computacional das adaptações dos algoritmos das Seções 2.2.2, 2.3.1 e 2.4.1 aos modelos mistos, e de todos os demais procedimentos envolvidos na modelagem dos $\text{PFMM1}(p)$ e $\text{PFMM2}(p_1, p_2)$, foram desenvolvidas funções na linguagem R, fazendo vasto uso da função `lme` do pacote `nlme`.

4.3.2 Etapa 2: Seleção dos efeitos aleatórios

Depois de selecionada a estrutura média preliminar, o próximo passo no processo de construção do PFMM é selecionar os efeitos aleatórios plausíveis. Devido ao fato de os FPs poderem também influenciar na estrutura dos efeitos aleatórios, é necessário verificar quais parâmetros associados à variabilidade dos efeitos aleatórios são significativos para decidir se a inclusão de efeitos aleatórios relacionados às curvaturas indicadas pelos FPs é necessária.

Testes para inclusão dos efeitos aleatórios baseados nos testes de permutação propostos por Lee & Braun (2012), apresentados na Seção 3.4, são utilizados nos delineamentos amostrais os quais, além do intercepto, demandam a inclusão de efeitos aleatórios associados às componentes dos FPs. Uma vez que o grau máximo que estamos admitindo para o modelo PFMM é 2, a matriz de covariâncias dos efeitos aleatórios, \mathbf{D} , para a i -ésima unidade amostral tem dimensão máxima (3×3) , no caso de se ter uma única variável regressora contínua. Para a modelagem das covariâncias desses efeitos, ou seja, \mathbf{b} , consideramos a matriz não estruturada.

Por simplicidade, ilustraremos a seleção dos efeitos aleatórios considerando apenas uma variável regressora contínua, x , cuja melhor transformação FP2 é dada por $p_1 \neq p_2$. Portanto, o melhor modelo PFMM2 selecionado na etapa 1 é do tipo

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + b_{0i} + \epsilon_{ij}, \quad i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad (4.5)$$

com $b_{0i} \sim \mathcal{N}(0, \sigma_0^2)$ e $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

Para testar a inclusão do efeito aleatório b_{1i} , associado à componente $x_{ij}^{p_1}$ no modelo (4.5), o teste de interesse envolve as seguintes hipóteses

$$\begin{aligned} H_0 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + b_{0i} + \epsilon_{ij}, \\ H_1 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + b_{0i} + b_{1i} x_{ij}^{p_1} + \epsilon_{ij}. \end{aligned} \quad (4.6)$$

Para testar as hipóteses em (4.6) utilizamos a estatística de teste apresentada

em (3.33) para $k = 2$. Caso o resultado do teste em (4.6) seja significativo, seguimos para verificar a necessidade de inclusão do efeito aleatório, b_{2i} , associado à componente x_{ij}^{p2} , dado que já foram incluídos os efeitos aleatórios b_{0i} e b_{1i} . Neste caso, o teste de interesse envolve as seguintes hipóteses

$$\begin{aligned} H_0 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p1} + \beta_2 x_{ij}^{p2} + b_{0i} + b_{1i} x_{ij}^{p1} + \epsilon_{ij}, \\ H_1 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p1} + \beta_2 x_{ij}^{p2} + b_{0i} + b_{1i} x_{ij}^{p1} + b_{2i} x_{ij}^{p2} + \epsilon_{ij}. \end{aligned} \quad (4.7)$$

Para testar as hipóteses em (4.7), a estatística de teste é novamente a apresentada em (3.33), mas para $k = 3$.

Caso o resultado do teste em (4.6) seja não significativo, seguimos para verificar a necessidade de inclusão do efeito aleatório, b_{2i} , associado a componente x_{ij}^{p2} , dado que já foi incluído apenas o intercepto aleatório b_{0i} no modelo. Nesse caso, as hipóteses a serem testadas são

$$\begin{aligned} H_0 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p1} + \beta_2 x_{ij}^{p2} + b_{0i} + \epsilon_{ij}, \\ H_1 : y_{ij} &= \beta_0 + \beta_1 x_{ij}^{p1} + \beta_2 x_{ij}^{p2} + b_{0i} + b_{2i} x_{ij}^{p2} + \epsilon_{ij}. \end{aligned} \quad (4.8)$$

Novamente, utilizamos a estatística de teste apresentada em (3.33) para $k = 2$.

O uso do teste de permutação baseado na estatística (3.33), utilizado para testar as hipóteses (4.6), (4.7) e (4.8), pode ser aplicado pois, em cada caso, testa-se a inclusão de um efeito aleatório por vez. Na situação de múltiplos efeitos aleatórios, o teste de permutação apropriado é aquele baseado na estatística da razão de verossimilhanças restrita generalizada.

4.3.3 Etapa 3: Seleção de uma estrutura de covariância para os erros

Após a seleção dos efeitos fixos e dos efeitos aleatórios, que modelam características populacionais e individuais, respectivamente, para o modelo PFMM, a variação restante nas respostas observadas deve-se ao erro aleatório.

Nesta etapa, busca-se avaliar a necessidade de uma estrutura diferente de $\sigma^2 \mathbf{I}$

para a matriz \mathbf{R}_i . Exemplos de estruturas de variância, ou funções de variância, e de correlação são apresentados nas Seções 3.3.1 e 3.3.2.

Para propor modelos para a função de variância e/ou estrutura de correlação para o modelo PFMM utilizamos a estratégia de análise apresentada por Singer et al. (2018), que incorporam alguns itens à proposta de Rocha (2004). A estratégia dos autores é definida conforme o tipo do estudo, longitudinal ou não longitudinal. A Seção 3.3.3 apresenta, em detalhes, todos os passos para a seleção da estrutura de covariância para cada um dos estudos.

Nesta etapa faremos uso de técnicas de diagnóstico para avaliar a qualidade do ajuste, conforme descrito na Seção 3.5.

4.3.4 Etapa 4: Redução do modelo

Na última etapa, selecionada a estrutura de covariância do modelo a partir das etapas 2 e 3, retornamos à etapa 1 e verificamos se a estrutura média, que foi selecionada sob o modelo de independência condicional homocedástico com apenas intercepto aleatório, se mantém. Espera-se, entretanto, que não haja grandes mudanças na estrutura média.

Ao final do processo, devemos realizar a análise de diagnóstico do modelo final por meio da construção de gráficos, conforme descrito na Seção 3.5.

5 APLICAÇÕES

5.1 Introdução

Neste capítulo apresentamos três aplicações da metodologia de modelagem proposta no Capítulo 4. As duas primeiras são provenientes de estudos com estrutura longitudinal, uma possui apenas uma variável regressora contínua e a outra possui duas variáveis regressoras, uma contínua e uma binária. A terceira aplicação apresenta uma estrutura de dados hierárquica com dois níveis de variação, em que temos setores censitários dentro de localidades, contando com doze variáveis regressoras contínuas e uma binária.

É importante ressaltar que a suposição padrão de observações independentes é violada nas três aplicações por causa da dependência entre as observações dentro da mesma unidade ou grupo. Métodos multinível e longitudinal discutidos nessas aplicações exploram a riqueza dos dados.

Para as análises dos dados das três aplicações seguimos a estratégia de modelagem descrita na Seção 4.3. As aplicações 1, 2 e 3 são apresentadas em detalhes, bem como os resultados das análises conduzidas, nas Seções 5.2, 5.3 e 5.4, respectivamente.

5.2 Aplicação 1: Concentração da bilirrubina em recém-nascidos nos primeiros doze dias de vida

O conjunto de dados longitudinais utilizado nessa aplicação foi analisado previamente em Rocha (2004) e é oriundo de estudo realizado na Escola Paulista de Medicina. Nessa pesquisa 89 recém-nascidos (RN) saudáveis em aleitamento ma-

terno foram avaliados. A variável de interesse é a concentração de bilirrubina (μ mol/L) que foi medida nas idades de 1, 2, 3, 4, 5, 6, 8, 10 e 12 dias de vida de todos os RNs.

O perfil médio da concentração de bilirrubina, Figura 5.1, sugere um polinômio para descrever a relação entre a concentração de bilirrubina e a idade do RN. O gráfico da Figura 5.2 mostra os perfis individuais cujo comportamento apresenta respostas iniciais e curvaturas diferentes ao longo das condições de avaliação (idades), sugerindo heterocedasticidade e a inclusão de interceptos e ou curvaturas como possíveis efeitos aleatórios, ambos associados à unidade amostral.

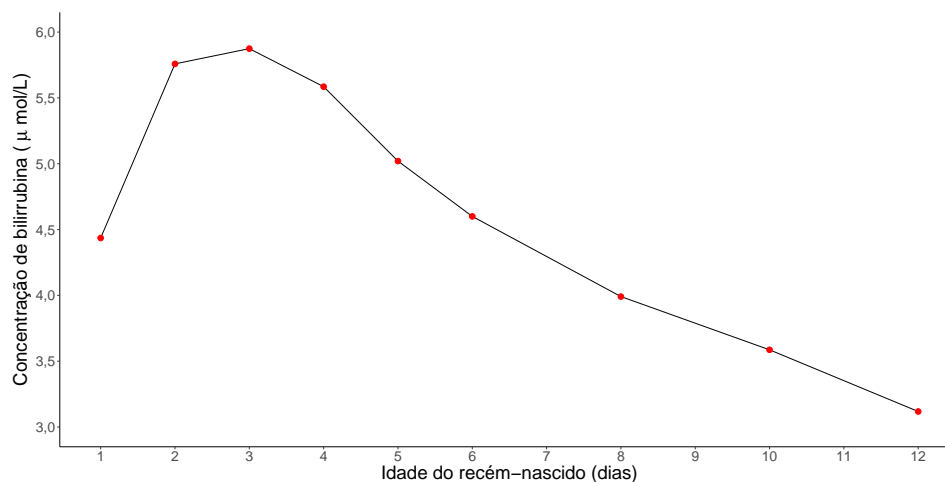


Figura 5.1. Perfil médio.

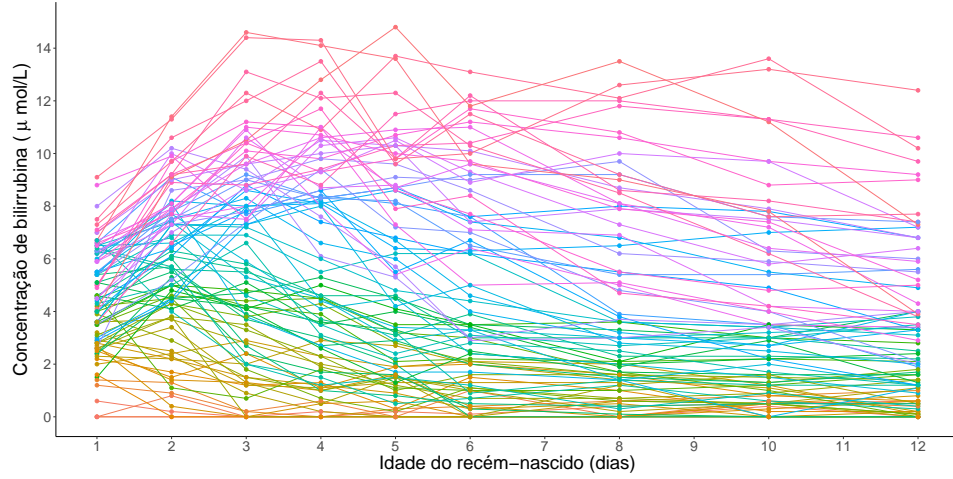


Figura 5.2. Perfis individuais.

5.2.1 Modelo PFMM

5.2.1.1 Etapa 1: Seleção de uma estrutura média preliminar

Foram ajustados os modelos PFMM1 e PFMM2 para representar a concentração de bilirrubina y_{ij} para a j -ésima idade x_{ij} para o i -ésimo RN, expressos por

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^p + b_{0i} + \epsilon_{ij}, \quad (5.1)$$

e

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + b_{0i} + \epsilon_{ij}, \quad (5.2)$$

respectivamente, com $i = 1, \dots, 89, j = 1, \dots, 9, b_{0i} \sim \mathcal{N}(0, \sigma_b^2)$ e $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Os parâmetros das distribuições dos termos aleatórios de (5.1) e (5.2) não necessariamente são iguais, mas prezando-se não carregar mais a notação, essa distinção não será explicitada. As potências p, p_1, p_2 foram selecionadas do conjunto S .

Os modelos PFMM1(2) e PFMM2(-0,5; -0,5) foram os que apresentaram o menor valor para a componente do desvio, conforme ilustrado nas Figuras 5.3 e 5.4, respectivamente.

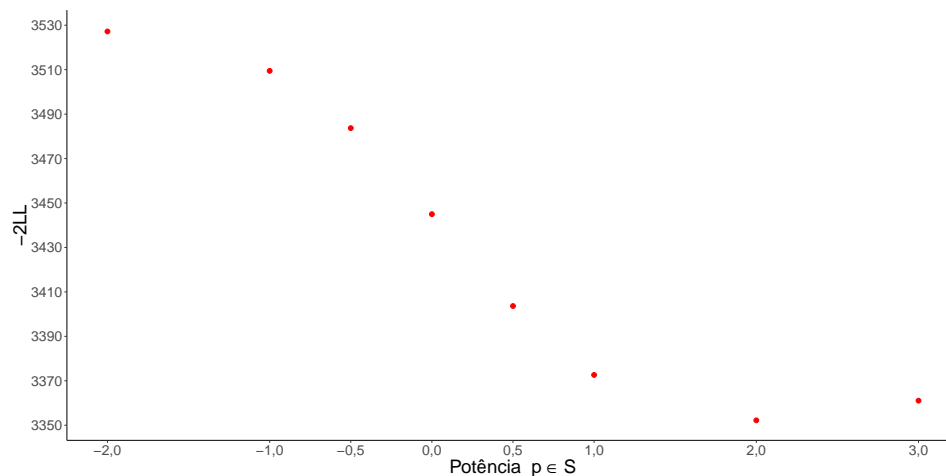


Figura 5.3. Gráfico da componente do desvio em função da potência $p \in S$ para o ajuste do modelo em (5.1).

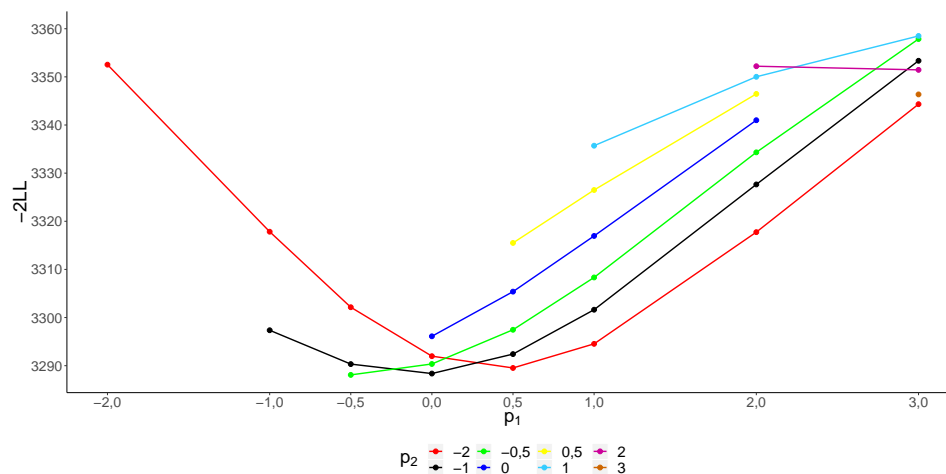


Figura 5.4. Gráfico da componente do desvio em função das potências $(p_1, p_2) \in S^2$ para o ajuste do modelo em (5.2).

Os passos do FSP e os resultados de cada teste são apresentados na Tabela 5.1 que mostra evidência favorável ao modelo $PFMM2(-0,5; -0,5)$, uma vez que (1) $PFMM2$ *versus* o modelo nulo é significativo, (2) $PFMM2$ *versus* modelo linear misto é significativo, e (3) $PFMM2$ *versus* $PFMM1$ também é significativo.

Tabela 5.1. Aplicação do procedimento FSP para selecionar o melhor modelo PFMM para dados da bilirrubina

Modelo	Graus de liberdade	-2LL	Potências	Comparação	T_{RVG}	P
PFMM2	7	3288,088	-0,5; -0,5	PFMM2 vs Nulo	240,666	< 0,001
PFMM1	5	3352,211	2	PFMM2 vs MLM	84,552	< 0,001
MLM	4	3372,640	1	PFMM2 vs PFMM1	64,123	< 0,001
Nulo	3	3528,754	-			

Portanto, o modelo selecionado tem grau 2 ($m = 2$) com a melhor transformação para as potências $(-0,5; -0,5)$, cuja expressão do preditor linear, ajustado por máxima verossimilhança, é dada por

$$\hat{\varphi}_2(x, \mathbf{p}) = -7,86(0,85) + 12,28(0,73)x_{ij}^{-0,5} + 10,32(0,69)x_{ij}^{-0,5} \ln(x_{ij}), \quad (5.3)$$

em que os valores entre parênteses são os erros padrão das estimativas dos coeficientes de regressão. Os valores de AIC , BIC e log-verossimilhança foram 3298,23, 3321,64 e $-1644,12$, respectivamente.

Na Figura 5.5 apresentamos o gráfico do perfil médio dos dados observados (pontos) com a sobreposição das curvas ajustadas pelos valores estimados do modelo (5.3) (PFMM2) e estimados de um modelo polinomial cúbico (PCMM). Pode-se observar que a estrutura média do modelo PFMM2 em (5.3) parece acomodar melhor os pontos observados. Espera-se uma assíntota para a curva que rege a concentração de bilirrubina.

Modelamos a estrutura média do modelo PFMM, mas temos interesse em avaliar o nível de heterogeneidade entre os RNs para verificar a necessidade de inclusão de efeitos aleatórios para as componentes dos FPs.

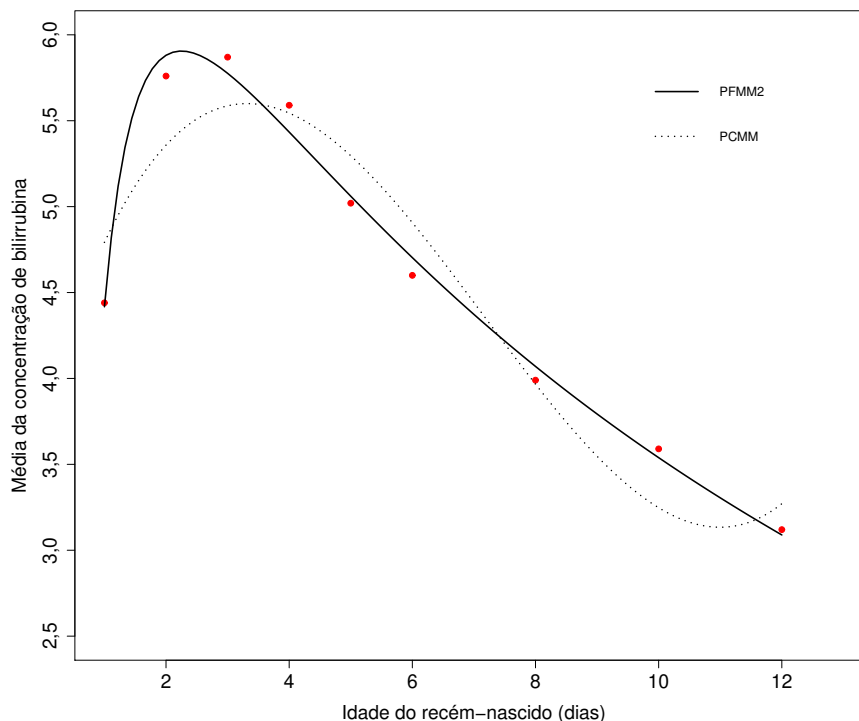


Figura 5.5. Gráfico do perfil médio dos dados observados e curvas ajustadas pelos valores estimados do modelo PFMM2 (5.3) e modelo polinomial cúbico com efeitos mistos (PCMM).

5.2.1.2 Etapa 2: Seleção dos efeitos aleatórios

Para testar se apenas uma ou ambas as componentes dos FPs selecionadas na etapa 1, $x_{ij}^{-0,5}$ e $x_{ij}^{-0,5} \ln(x_{ij})$, devem ser incluídas aos perfis individuais foram realizados o teste de permutação baseado no BLUP e o teste assintótico baseado na estatística da RVG.

A Tabela 5.2 apresenta os resultados de ambos os testes. Dado o intercepto aleatório no modelo, o efeito aleatório para a componente $x_{ij}^{-0,5}$ foi significativo. Na sequência, foi testada a inclusão do efeito aleatório para a componente $x_{ij}^{-0,5} \ln(x_{ij})$, com b_{0i} e $b_{1i}x_{ij}^{-0,5}$ no modelo. Novamente, o resultado do teste foi significativo.

Tabela 5.2. Resultados dos testes de permutação e assintótico para inclusão de efeitos aleatórios específicos ao modelar a concentração de bilirrubina em RNs

Teste	Estatística observada		P	
	BLUP	TRVR	Permutação	Assintótico
(1) vs (2)	1252,326	202,484	< 0,0001	< 0,0001
(2) vs (3)	6062,449	245,709	< 0,0001	< 0,0001

$$(1): y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0,5} + \beta_2 x_{ij}^{-0,5} \ln(x_{ij}) + b_{0i} + \epsilon_{ij}$$

$$(2): y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0,5} + \beta_2 x_{ij}^{-0,5} \ln(x_{ij}) + b_{0i} + b_{1i} x_{ij}^{-0,5} + \epsilon_{ij}$$

$$(3): y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0,5} + \beta_2 x_{ij}^{-0,5} \ln(x_{ij}) + b_{0i} + b_{1i} x_{ij}^{-0,5} + b_{2i} x_{ij}^{-0,5} \ln(x_{ij}) + \epsilon_{ij}.$$

Portanto, temos evidências de que coeficientes relacionados aos termos $x_{ij}^{-0,5}$ e $x_{ij}^{-0,5} \ln(x_{ij})$ variam entre os RNs e devem ser incluídos aos perfis individuais. Damos prosseguimento a análise com o seguinte modelo

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) x_{ij}^{-0,5} + (\beta_2 + b_{2i}) x_{ij}^{-0,5} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.4)$$

em que

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2),$$

com

$$\mathbf{D} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{bmatrix}. \quad (5.5)$$

Na Tabela 5.3 encontram-se as estimativas de máxima verossimilhança restrita para os parâmetros do modelo (5.4).

Tabela 5.3. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo em (5.4)

Parâmetro	Estimativa	Erro padrão
β_0	-7,86	1,00
β_1	12,28	1,08
β_2	10,32	1,03
σ^2	0,89	
σ_0^2	68,83	
σ_1^2	86,10	
σ_2^2	78,73	
σ_{01}	-0,99	
σ_{02}	-0,87	
σ_{12}	0,89	
<i>AIC</i>	2860,04	
<i>BIC</i>	2906,86	
Log-verossimilhança	-1420,02	

5.2.1.3 Etapa 3: Seleção de uma estrutura de covariâncias para os erros

Na etapa 2 finalizamos selecionando o modelo expresso em (5.4) cuja estrutura de covariâncias para os erros é dada por $\mathcal{R}_i = \sigma^2 \mathbf{I}_{n_i}$, indicando que as n_i observações associadas à i -ésima unidade amostral são condicionalmente independentes dado \mathbf{b}_i .

A matriz (5.6) apresenta as variâncias (negrito) e correlações amostrais. Sem os elementos redundantes, pode-se notar que: i) as variâncias crescem e depois decrescem ao longo dos dias após o nascimento, indicando heterocedasticidade; ii) as correlações diminuem ao longo dos dias após o nascimento, indicando uma possível correlação serial, e iii) as correlações em cada sub-diagonal apresentam um comportamento crescente.

A Figura 5.6 apresenta o gráfico das linhas da matriz (5.6), sugerindo a presença de correlação serial, detectada pelo decréscimo dos perfis à medida que a distância entre as observações aumenta, confirmando a análise baseada na matriz (5.6).

A Figura 5.7 apresenta os gráficos do variograma amostral que evidenciam

também a presença de correlação serial nos dados. A linha horizontal apresentada na Figura 5.7 é a estimativa da variância residual baseada no variograma, no caso $\hat{\sigma}^2 = 13,04$.

	dia1	dia2	dia3	dia4	dia5	dia6	dia8	dia10	dia12
dia1	3,98	0,82	0,71	0,63	0,55	0,52	0,51	0,50	0,48
dia2		8,50	0,90	0,86	0,79	0,76	0,72	0,70	0,68
dia3			14,92	0,95	0,91	0,88	0,85	0,82	0,78
dia4				16,97	0,95	0,93	0,88	0,85	0,81
dia5					16,20	0,96	0,94	0,91	0,85
dia6						15,25	0,96	0,93	0,88
dia8							13,70	0,98	0,94
dia10								11,50	0,96
dia12									8,85

(5.6)

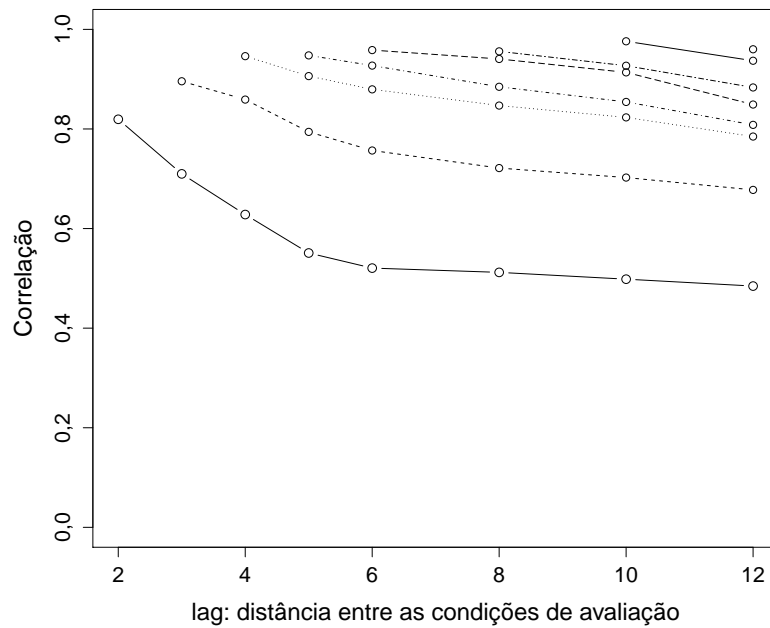


Figura 5.6. Correlações em função da defasagem.

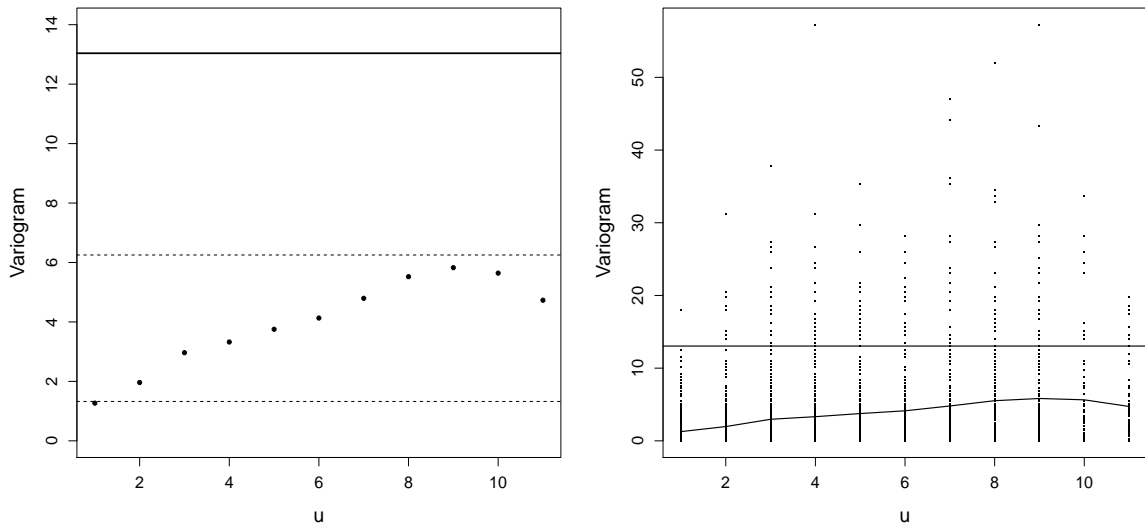


Figura 5.7. Variograma amostral.

Com base nas indicações descritivas de correlação serial e heterocedasticidade nos dados da bilirrubina, seguimos a modelagem modificando a estrutura de covariâncias do erro do modelo (5.4) fazendo combinações de três funções de variância, sendo as funções exponencial, potência e heterogênea, isto é, com diferentes variâncias por idade, e três estruturas de correlação serial, AR1, ARMA(1,1) e Não Estruturada (NE). Foram ajustados nove modelos.

A Tabela 5.4 apresenta os valores de $-2 \times \log$ -verossimilhança restrita ($-2LL_R$), AIC e BIC desses nove modelos ajustados. Utilizamos os critérios AIC e BIC para comparação dos nove modelos, dado que, possivelmente, os modelos não são encaixados, e pensando em penalizar a verossimilhança em relação ao número de parâmetros. O modelo (2c), com diferentes variâncias para cada nível da variável regressora idade e estrutura de correlação serial ARMA(1,1), é o que melhor representa a matriz de covariâncias dos erros aleatórios, com $AIC = 2676,67$.

Tabela 5.4. Estatísticas de ajuste do modelo (5.4) com modificações na estrutura de covariâncias do erro para três diferentes funções de variância e estruturas de correlação

Modelo	Função de variância	Estrutura de correlação	Número de parâmetros	$-2LL_R$	AIC	BIC
(1a)	Exponencial	AR1	14	2686,71	2710,72	2766,90
(1b)	Potência		14	2702,87	2726,87	2783,06
(1c)	Heterogênea		21	2645,99	2683,99	2772,95
(2a)	Exponencial	ARMA(1,1)	15	2684,86	2710,86	2771,73
(2b)	Potência		15	2701,43	2727,43	2788,30
(2c)	Heterogênea		22	2636,67	2676,67	2770,32
(3a)	Exponencial	NE	49	2599,78	2693,78	2913,84
(3b)	Potência		49	2601,14	2695,14	2915,20
(3c)	Heterogênea		56	2572,53	2680,53	2933,37

A Figura 5.8 apresenta os dados observados da concentração de bilirrubina em função da idade, por recém-nascido, e as curvas ajustadas pelos modelos (5.4) e (2c). Apesar da indicação da presença de correlação serial nos dados, a Figura 5.8 aponta que o modelo (5.4), com uma estrutura de covariância homocedástica, parece ajustar um pouco melhor. O modelo (2c) apresenta $AIC = 2676,67$ e o modelo (5.4) $AIC = 2860,04$.

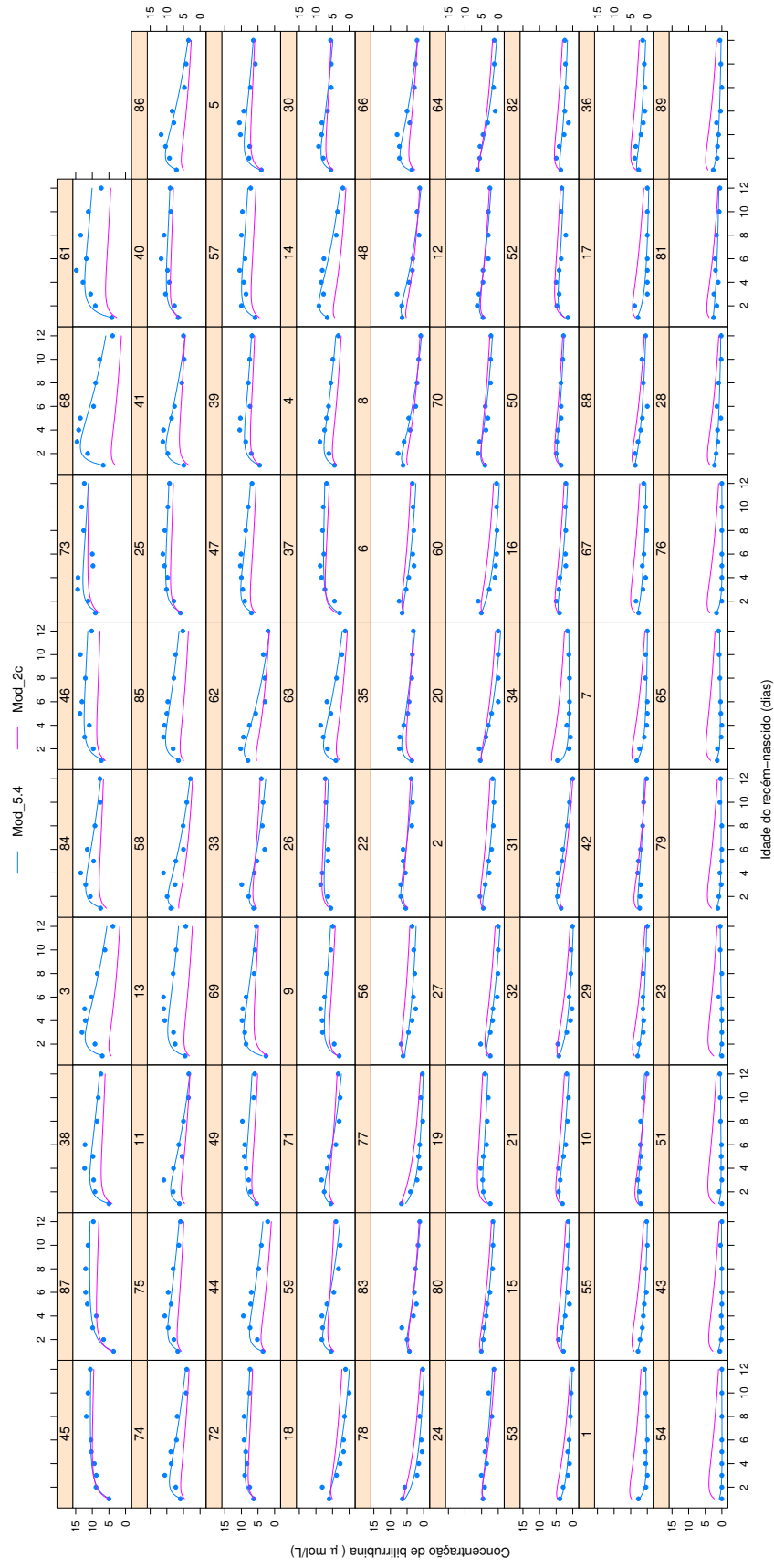


Figura 5.8. Dados observados (pontos) e curvas de predição para os modelos (5.4) e (2c).

Consideramos então duas modificações na estrutura de covariâncias dos erros definida em (2c): (i) permitimos que os erros dentro de grupos de idades com variâncias amostrais similares tenham suas próprias variações e, para tanto, utilizamos a matriz de variâncias amostral dada em (5.6). Nesse sentido, agrupamos a variável regressora idade em quatro grupos distintos tal que, RNs com idade de 1 dia foram alocadas no grupo 1; com idades de 2 e 12 dias no grupo 2; com idades de 3 a 6 dias no grupo 3; e com idades de 8 e 10 dias no grupo 4; (ii) excluímos a estrutura de correlação, ARMA(1,1), visto que parte da variabilidade dos dados se encontra modelada na estrutura dos efeitos aleatórios.

Assumindo variância heterocedástica para cada grupo de idade, damos prosseguimento a análise com o seguinte modelo

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) x_{ij}^{-0,5} + (\beta_2 + b_{2i}) x_{ij}^{-0,5} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.7)$$

com

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 \delta_{sij}^2),$$

em que $i = 1, \dots, 89$, $j = 1, 2, 3, 4, 5, 6, 8, 10, 12$, $s = 1, \dots, 4$, tal que $\delta_1 = 1$ e $\delta_s > 0$, $s = 2, 3, 4$, e a matriz \mathbf{D} não estruturada.

O teste da RVRG entre os modelos em (5.4) e (5.7) apresentou um aumento significativo na log-verossimilhança do modelo (5.7), $P < 0,001$. Na Tabela 5.5 encontram-se as estimativas de máxima verossimilhança restrita para os parâmetros do modelo em (5.7).

Uma avaliação da versão heteroscedástica do modelo PFMM2 em (5.7) é fornecida pela Figura 5.9. Podemos observar que os valores preditos se ajustam muito bem às concentrações de bilirrubina observadas.

A fim de avaliar a qualidade do ajuste do modelo PFMM2 dado em (5.7), análises gráficas dos resíduos são apresentadas nas Figuras 5.10–5.15.

A Figura 5.10 apresenta o gráfico das componentes dos resíduos condicionais padronizados em função dos valores ajustados e verifica-se um padrão razoavelmente homogêneo de variabilidade para esses resíduos. Analisando a Figura 5.11, observa-se um padrão condizente com a hipótese de linearidade dos efeitos fixos embora, em valores absolutos, os resíduos abaixo da reta de referência no zero são menores do que os acima da reta.

Pela Figura 5.12, não se percebe afastamento importante da normalidade por parte do resíduo com confundimento mínimo, indicando ser plausível a suposição de normalidade do erro condicional.

Tabela 5.5. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.7)

Parâmetro	Estimativa	Erro padrão
β_0	-7,90	1,06
β_1	12,33	1,14
β_2	10,34	1,05
σ^2	4,46e-11	
σ_0^2	82,92	
σ_1^2	99,29	
σ_2^2	86,12	
σ_{01}	-0,98	
σ_{02}	-0,88	
σ_{12}	0,90	
δ_2	178417,52	
δ_3	143103,73	
δ_4	83994,17	
<i>AIC</i>	2802,98	
<i>BIC</i>	2863,84	
Log-verossimilhança	-1388,49	

O gráfico para a estrutura de covariância destaca os RNs 61 e 73 com valores acima da faixa prevista (Figura 5.13).

O gráfico da distância de Mahalanobis em função das unidades amostrais des-

taca algumas unidades como discrepantes (Figura 5.14). No entanto, esses valores discrepantes parecem não afetar a normalidade dos efeitos aleatórios (Figura 5.15).

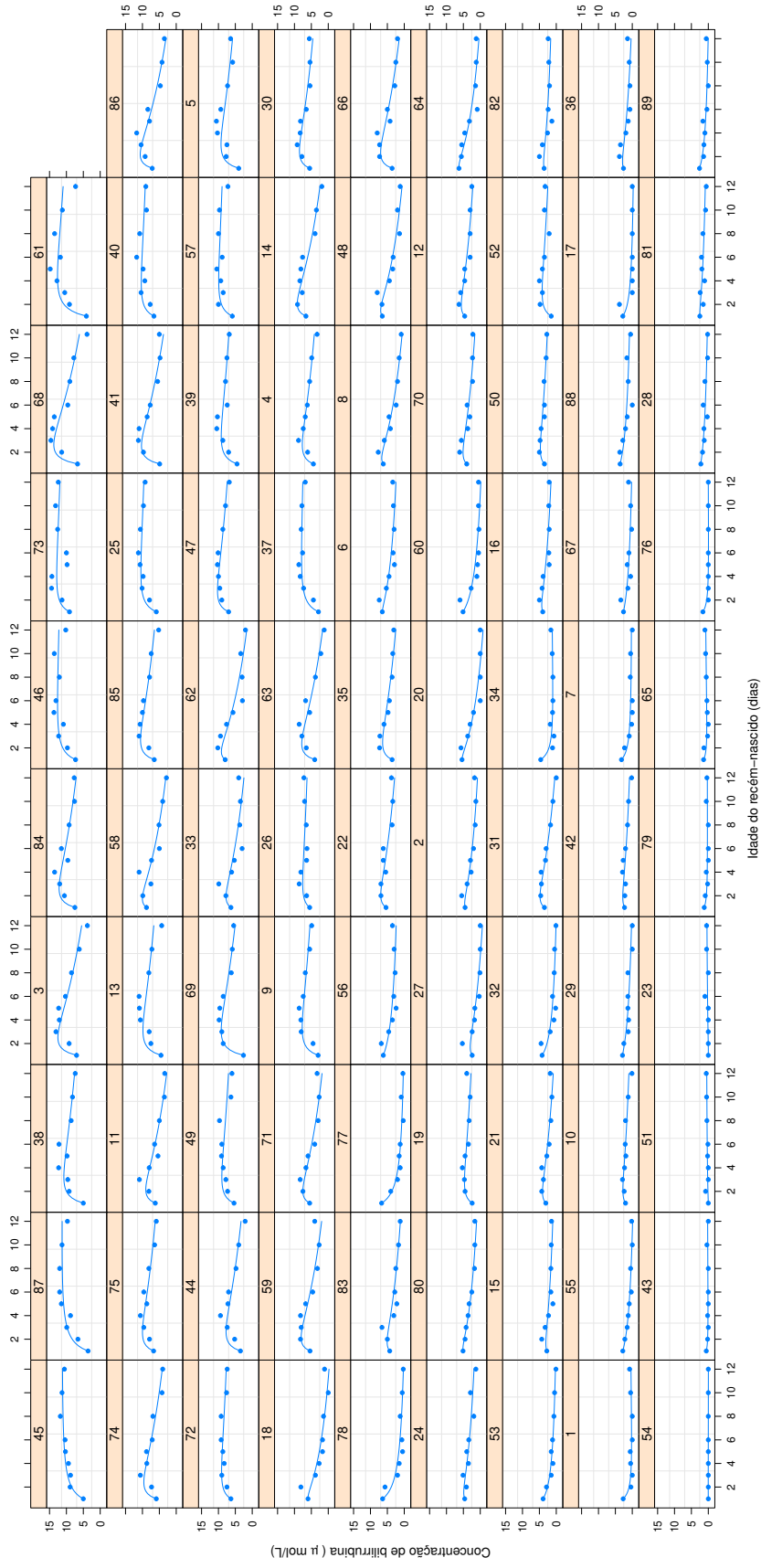


Figura 5.9. Dados observados (pontos) e curvas de predição para o modelo (5.7).

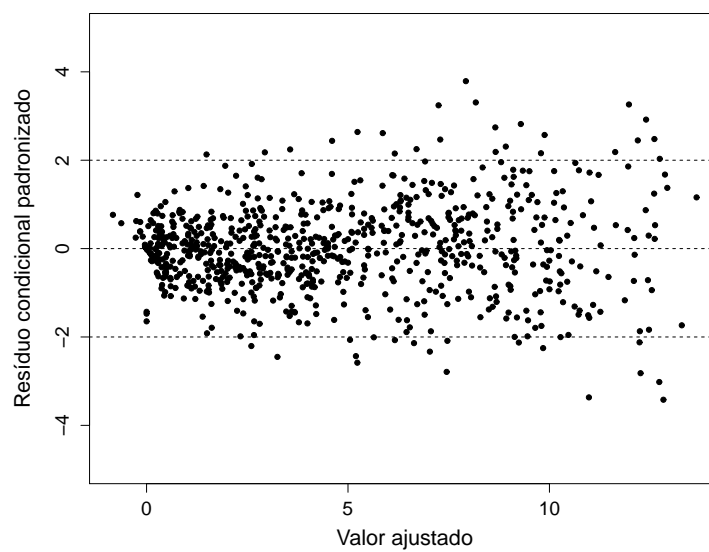


Figura 5.10. Gráfico do resíduo condicional padronizado *versus* valor ajustado para o modelo em (5.7).

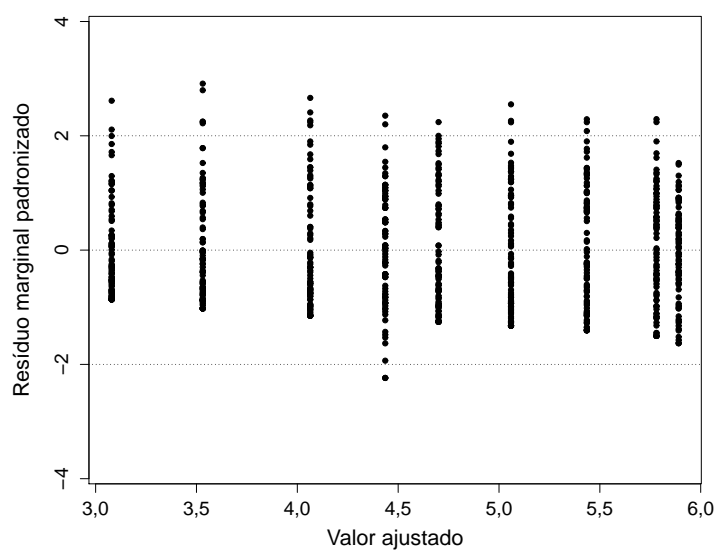


Figura 5.11. Gráfico do resíduo marginal padronizado *versus* valor ajustado para o modelo em (5.7).

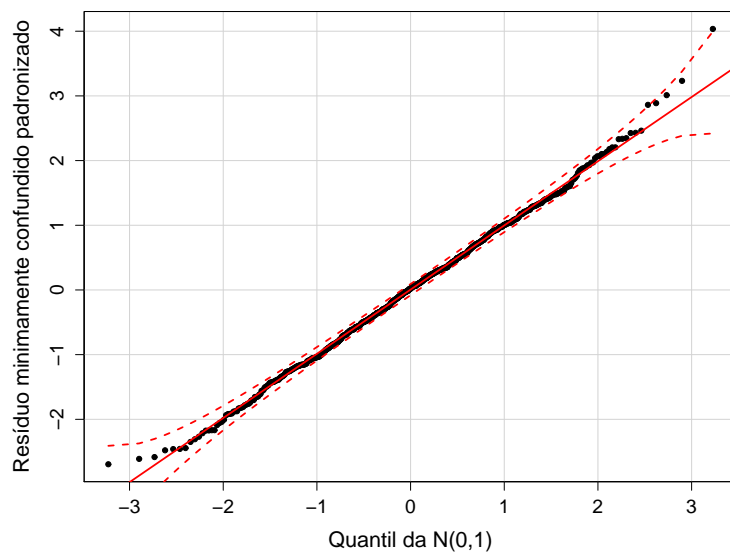


Figura 5.12. Gráfico do envelope simulado com 95% para resíduos condicionais minimamente confundidos para o modelo em (5.7).

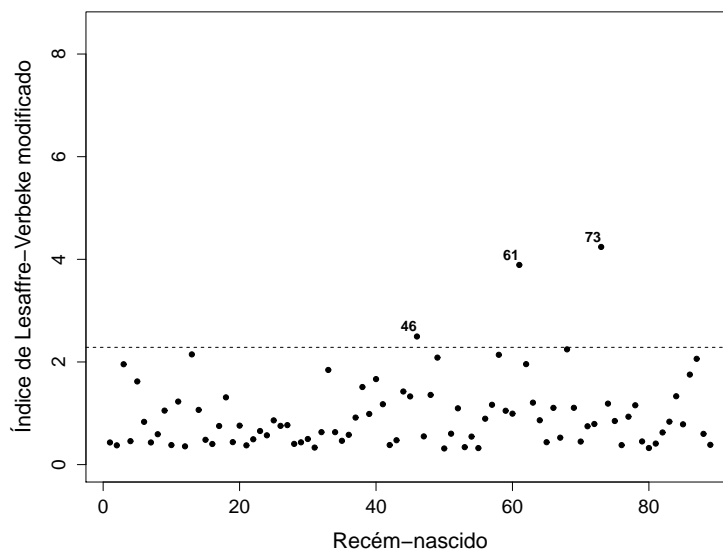


Figura 5.13. Gráfico do índice de Lesaffre-Verbeke modificado *versus* índice do recém-nascido para o modelo (5.7).

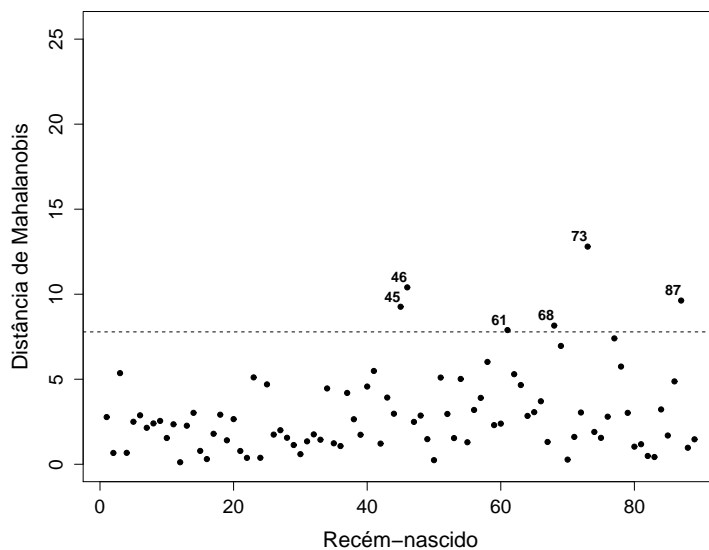


Figura 5.14. Gráfico baseado na distância de Mahalanobis *versus* índice do recém-nascido para o modelo em (5.7).

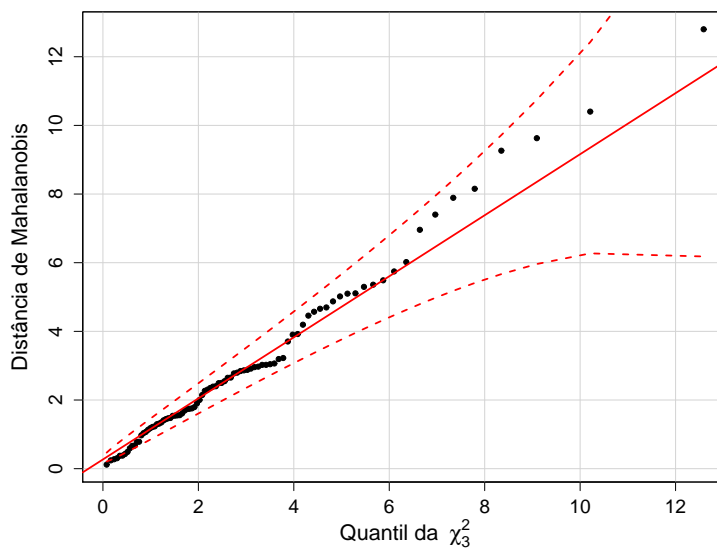


Figura 5.15. Gráfico QQ baseado na distância de Mahalanobis correspondente ao modelo em (5.7).

5.2.1.4 Etapa 4: Redução do modelo

Após varrer novamente o conjunto S , mas agora incorporando os efeitos aleatórios selecionados na etapa 2 e estrutura de covariância para erro selecionada na etapa 3, o modelo PFMM com potências $p_1 = -0,5$ e $p_2 = 0$ foi o que apresentou menor valor para a componente do desvio ($-2LL = 2770,62$), conforme ilustrado na Figura 5.16. Nota-se que para as potências $(0; 0)$ e $(-0,5; -0,5)$ os valores das componente do desvio, 2772,07 e 2777,68, respectivamente, são bem próximos das potências $(-0,5; 0)$.

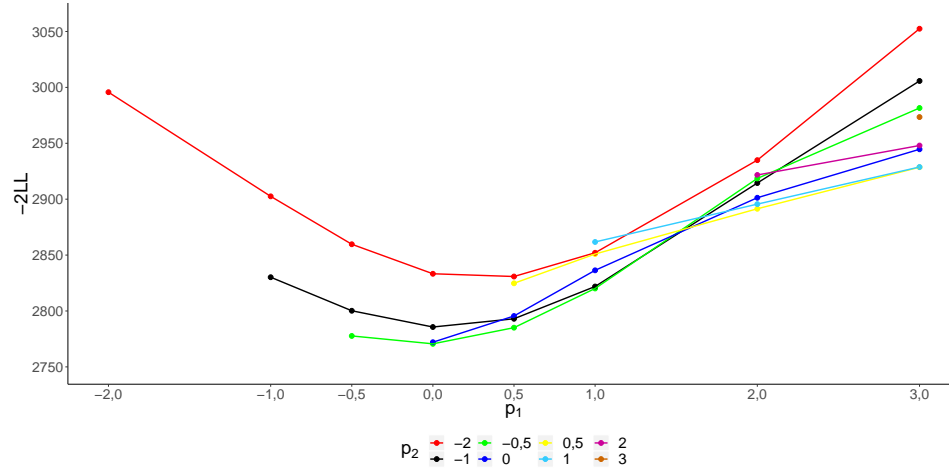


Figura 5.16. Gráfico da componente do desvio para 36 modelos $PFMM2(p_1, p_2)$ com efeitos aleatórios e estrutura de covariância do erro selecionados nas etapas 2 e 3, respectivamente.

Portanto, o modelo selecionado na etapa 4 é um $PFMM2(-0,5; 0)$, dado por

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{-0,5} + \beta_2 \ln(x_{ij}) + b_{0i} + b_{1i} x_{ij}^{-0,5} + b_{2i} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.8)$$

$$i = 1, \dots, 89; j = 1, 2, 3, 4, 5, 6, 8, 10, 12,$$

em que

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 \delta_{s_{ij}}^2),$$

sendo a matriz \mathbf{D} não estruturada, $s = 1, \dots, 4$, tal que $\delta_1 = 1$ e $\delta_s > 0$, $s = 2, 3, 4$.

A mudança na transformação $FP2(-0,5; -0,5)$ para $FP2(-0,5; 0)$ simplificou o modelo com a retirada da variável regressora $x^{-0,5}$ na componente β_2 , além de reduzir o AIC para 2797,12. Na Tabela 5.6 encontram-se as estimativas de máxima verossimilhança restrita para os parâmetros do modelo (5.8).

Tabela 5.6. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.8)

Parâmetro	Estimativa	Erro padrão
β_0	22,26	2,19
β_1	-17,82	2,10
β_2	-5,70	0,58
σ^2	8,76e-11	
σ_0^2	390,08	
σ_1^2	355,86	
σ_2^2	26,74	
σ_{01}	-1,00	
σ_{02}	-0,98	
σ_{12}	0,97	
δ_2	122722,50	
δ_3	102251,09	
δ_4	61074,28	
AIC	2797,12	
BIC	2857,99	
Log-verossimilhança	-1385,56	

Na Figura 5.17 apresentamos o gráfico do perfil médio dos dados observados (pontos) com a sobreposição da curva ajustada pelos valores preditos do modelo (5.8). A ilustração inclui um período de extrapolação de dois dias além dos dados observados. Nota-se que a estrutura média do modelo (5.8) se acomoda bem aos dados.

A Figura 5.18 apresenta a taxa de variação da curva de crescimento do preditor linear do modelo em (5.8), em função da idade, também com período de extrapolação de dois dias. É possível notar que a taxa de variação dos níveis de bilirrubina começa

alta e cai abruptamente até cerca da idade de 4 dias, após a qual parece tender a se estabilizar. O pico médio dos níveis de bilirrubina ocorre, aproximadamente, na idade de 2,5 dias. As curvas individuais ajustadas pelo modelo em (5.8) são apresentadas na Figura 5.19.

A Figura 5.20 apresenta gráficos de diagnóstico do modelo em (5.8). Comparando estes gráficos aos gráficos de diagnóstico do modelo em (5.7) não se observa mudanças entre eles. Gráficos de influência para o modelo em (5.8) são apresentados nas Figuras 6.1– 6.7 do Anexo e não indicam observações ou unidades influentes.

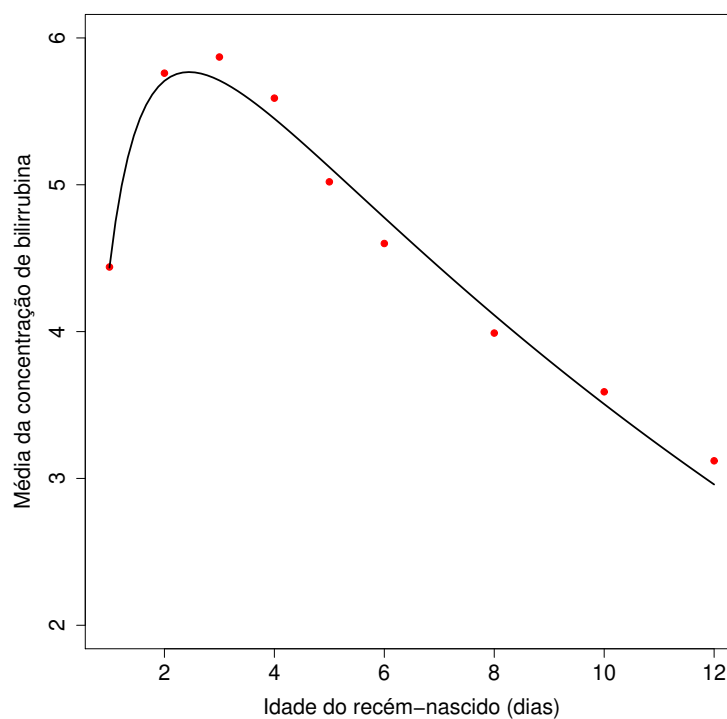


Figura 5.17. Gráfico do perfil médio dos dados observados e curva ajustada pelos valores preditos do modelo em (5.8).

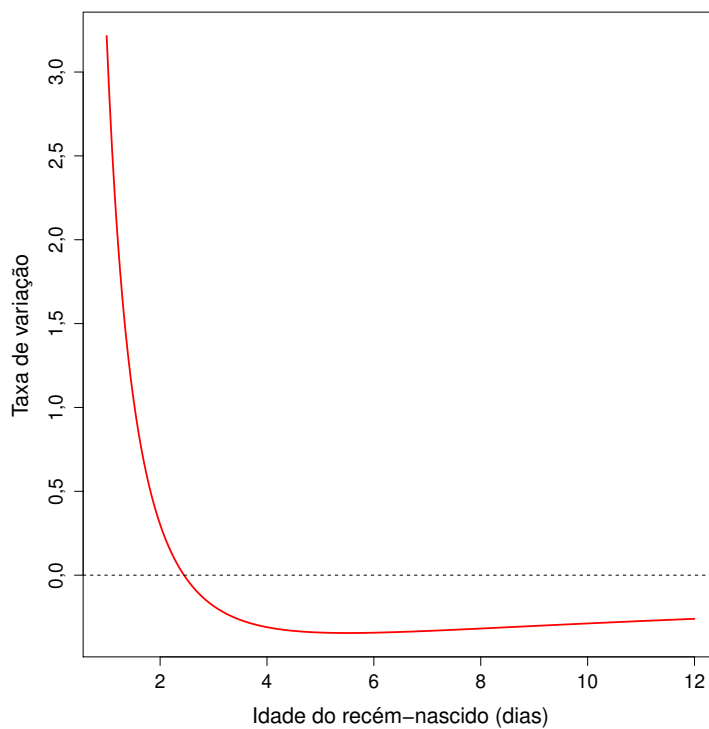


Figura 5.18. Gráfico da taxa de variação para a curva média marginal em (5.8) em função da idade.

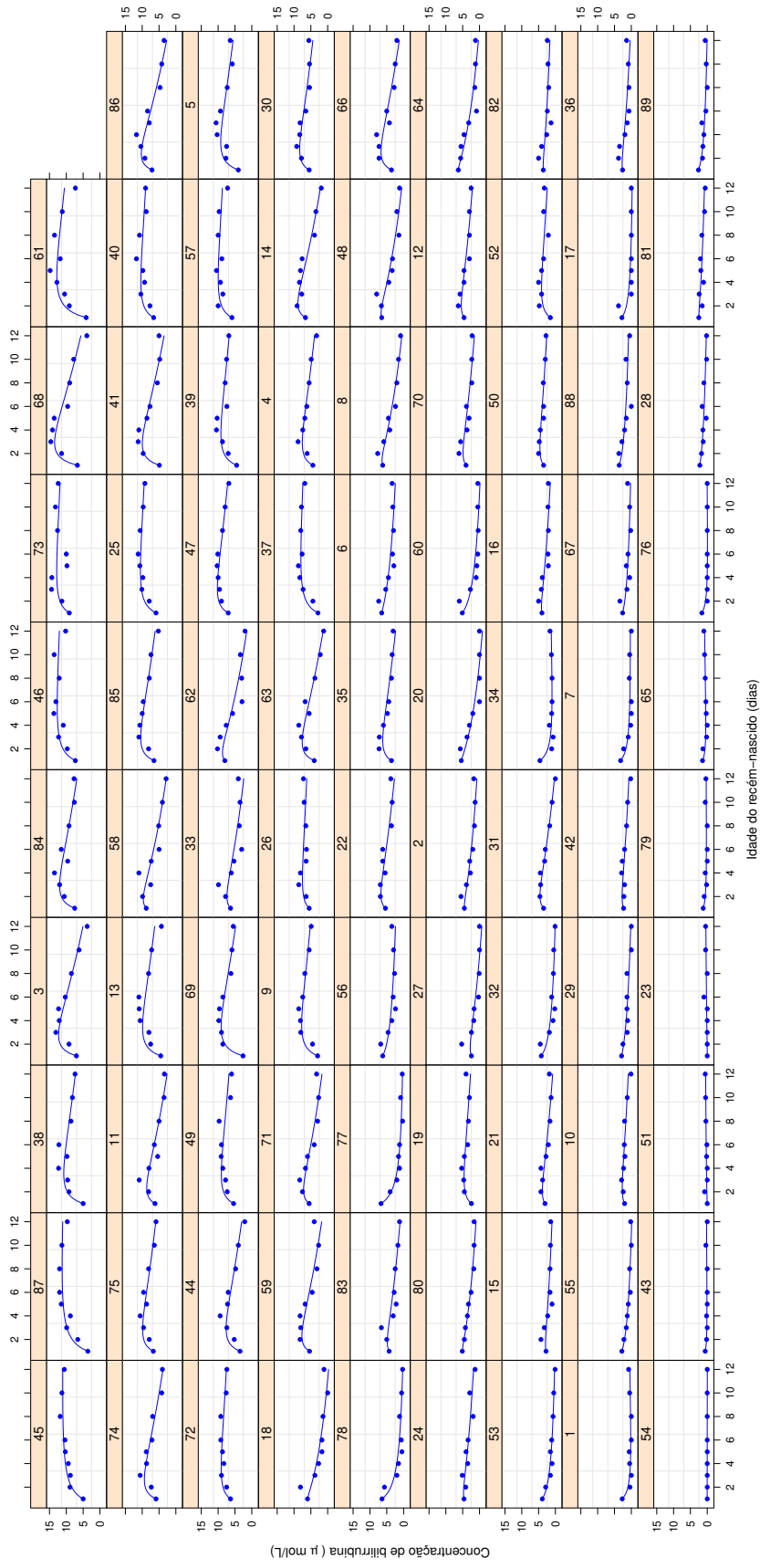


Figura 5.19. Dados observados (pontos) e curvas de predição para o modelo (5.8).

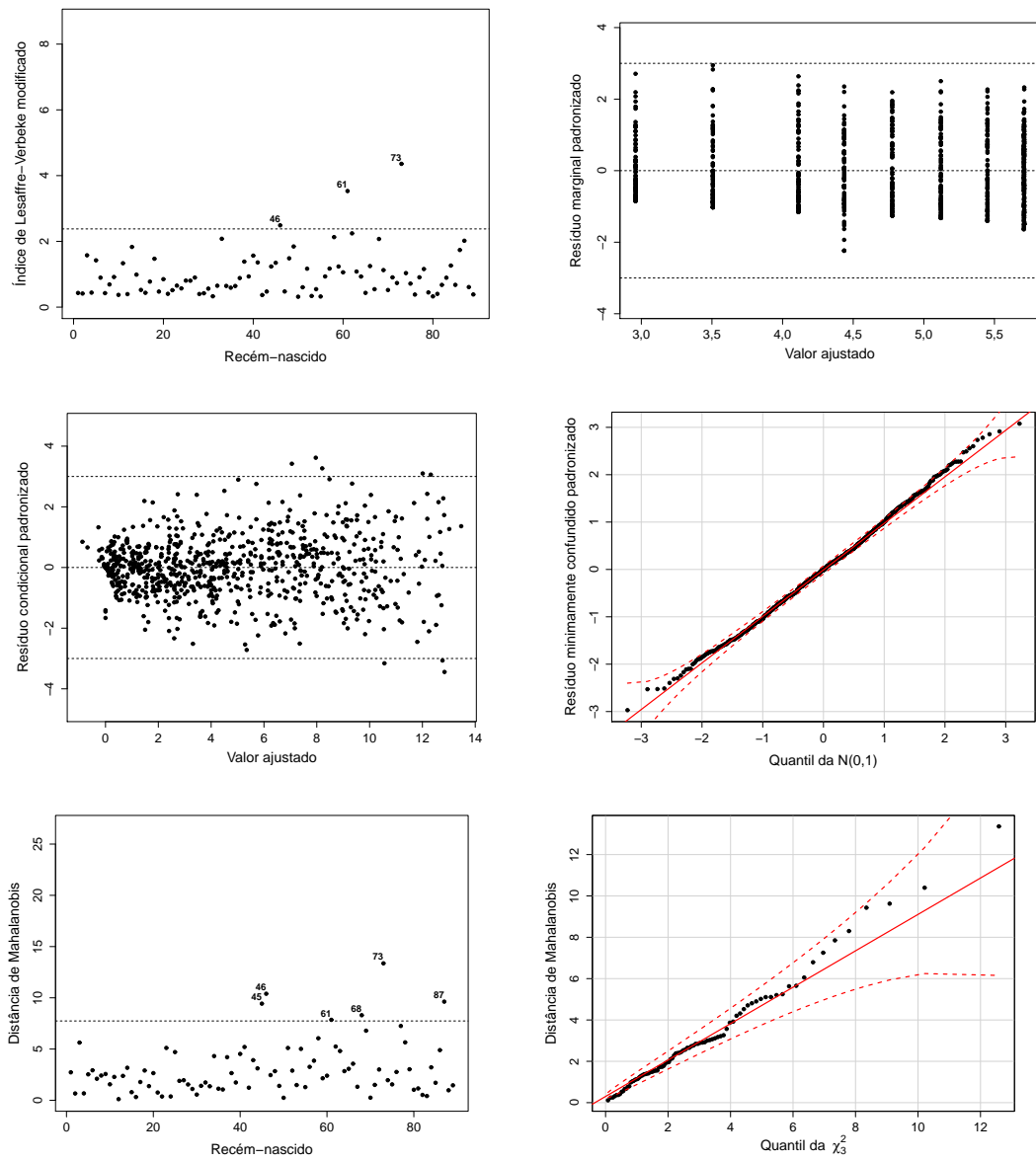


Figura 5.20. Gráficos de diagnóstico para o modelo em (5.8).

5.3 Aplicação 2: Dados da taxa de ultrafiltração de dialisadores de alto fluxo

5.3.1 Descrição do estudo

Os dados são provenientes de um estudo de Vonesh & Carter (1992) que avaliaram a taxa de ultrafiltração de dialisadores de alto fluxo utilizados no tratamento de pacientes com insuficiência renal crônica. Os autores verificaram que dialisadores de alto fluxo têm cinética de transporte de água caracterizada por uma relação não linear entre a taxa de ultrafiltração e a pressão transmembrana. Dessa forma, utilizaram um modelo não linear para representar a relação entre essas variáveis.

Littell et al. (1996) e Pinheiro & Bates (2000) apresentaram uma análise alternativa utilizando os modelos lineares mistos. Os dados encontram-se disponíveis no pacote `nlme` da linguagem R com a denominação `Dialyzer`.

O estudo envolveu 20 dialisadores de alto fluxo medidos em 7 níveis crescentes de pressão transmembrana (dmHg). Os dialisadores foram divididos em dois grupos e avaliados *in vitro* usando sangue bovino. O primeiro grupo consistiu de 10 dialisadores avaliados a uma taxa de fluxo de sangue de 200 dl/min, enquanto o segundo grupo de 10 dialisadores foi avaliado a uma taxa de fluxo de sangue de 300 dl/min.

O conjunto de dados contém, portanto, 140 observações e três variáveis. A variável resposta é a taxa de ultrafiltração (ml/hr), a pressão transmembrana é uma variável regressora contínua e a taxa de fluxo de sangue uma variável regressora binária. A natureza não linear da relação entre a taxa de ultrafiltração e a pressão transmembrana é apresentada na Figura 5.21.

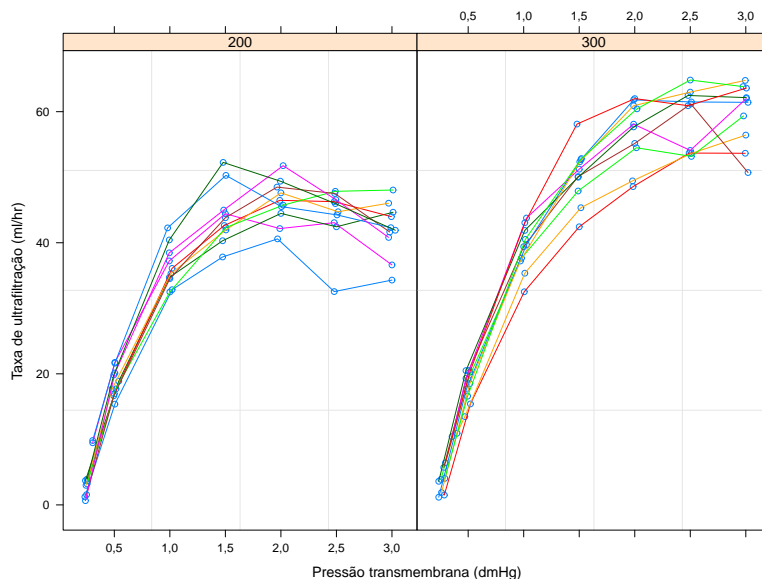


Figura 5.21. Perfis individuais dos dialisadores segundo a taxa de ultrafiltração (ml/hr) medida em 7 níveis crescentes de pressão transmembrana (dmHg) para as taxas de fluxo de sangue de 200 dl/min e 300 dl/min.

A forma dessas curvas sugere que a taxa de ultrafiltração aumenta com a pressão transmembrana até um máximo e que as maiores taxas de ultrafiltração são obtidas para dialisadores avaliados com fluxo de sangue de 300 dl/min. Deve-se considerar que medições feitas no mesmo dialisador sejam correlacionadas e é possível notar uma indicação de que a variabilidade nas taxas de ultrafiltração aumenta com a pressão transmembrana.

5.3.2 Modelo linear misto

O melhor modelo linear misto indicado nas análises de Pinheiro & Bates (2000), para representar a taxa de ultrafiltração y_{ij} para a j -ésima pressão transmembrana x_{ij} para o i -ésimo dialisador, é expresso por

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + \gamma_1 w_i + b_{1i}) x_{ij} + (\beta_2 + b_{2i}) x_{ij}^2 + \beta_3 x_{ij}^3 + \beta_4 x_{ij}^4 + \epsilon_{ij}, \quad i = 1, \dots, 20, \quad j = 1, \dots, 7, \quad (5.9)$$

em que

$$w_i = \begin{cases} 0, & \text{se QB} = 200 \text{ dl/min} \\ 1, & \text{se QB} = 300 \text{ dl/min}, \end{cases}$$

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim N_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim N(0, \sigma^2 x_{ij}^{2\delta}), \quad \delta \in \mathbb{R},$$

e

$$\mathbf{D} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{bmatrix}. \quad (5.10)$$

Os parâmetros $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_0$ e γ_1 são efeitos fixos e representam intercepto, efeito linear, quadrático, cúbico e quártico da pressão transmembrana, efeito principal do fluxo de sangue e efeito de interação entre fluxo de sangue e pressão transmembrana, respectivamente; \mathbf{b}_i é o vetor de efeitos aleatórios, cuja estrutura de covariância especificada para a matriz \mathbf{D} é definida positiva; ϵ_{ij} é o erro aleatório, cuja variância se altera com a pressão transmembrana, de acordo com o modelo de potência com parâmetro δ (estimado junto com os outros parâmetros no modelo), utilizado devido ao comportamento heteroscedástico observado nos resíduos do modelo misto com erros homocedásticos e sendo assumido que \mathbf{b}_i e ϵ_{ij} são independentes.

Iremos nos reportar ao modelo (5.9) como PCMM que seria um modelo com polinômios convencionais no contexto dos modelos mistos.

Para fins de discussão de um modelo alternativo que proporemos para este conjunto de dados, apresentamos na Tabela 5.7 as estimativas da máxima verossimilhança restrita dos parâmetros obtidas no ajuste do modelo (5.9).

Tabela 5.7. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.9)

Parâmetro	Estimativa	Erro padrão
β_0	-16,39	0,83
β_1	90,62	3,27
β_2	-49,69	4,23
β_3	12,47	2,15
β_4	-1,28	0,36
γ_0	-2,78	0,83
γ_1	7,03	0,79
σ^2	1,59	
σ_0^2	3,53	
σ_1^2	28,77	
σ_2^2	2,77	
σ_{01}	-0,52	
σ_{02}	0,37	
σ_{12}	-0,96	
δ	0,76	
<i>AIC</i>	648,96	
<i>BIC</i>	692,31	
Log-verossimilhança	-309,48	

5.3.3 Modelo PFMM

O diferencial na análise desse conjunto de dados em relação à Aplicação 1 está na investigação de termos de interação entre as variáveis regressoras, pressão transmembrana e fluxo de sangue, por meio do algoritmo MFPI, descrito na Subseção 2.4.1. A seguir, apresentamos os resultados do desenvolvimento das quatro etapas da estratégia proposta na Seção 4.3 para selecionar o melhor modelo PFMM utilizando o algoritmo MFPI.

5.3.3.1 Etapa 1: Seleção de uma estrutura média preliminar

Para seleção do modelo para a média da resposta foram ajustadas 36 funções FP2 para x_{ij} considerando o modelo PFMM2

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_2} + \gamma_0 w_i + b_{0i} + \epsilon_{ij}, \quad (5.11)$$

quando $p_1 \neq p_2$, e

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{p_1} + \beta_2 x_{ij}^{p_1} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij}, \quad (5.12)$$

quando $p_1 = p_2$, ambos com $b_{0i} \sim \mathcal{N}(0, \sigma_0^2)$ e $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

A Figura 5.22 apresenta os valores da componente do desvio para os modelos (5.11) e (5.12) em função da potência p_1 estratificada por p_2 . Observa-se que o melhor modelo PFMM2 tem potências (1, 1), com $-2LL = 837,28$, sendo expresso por

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij}. \quad (5.13)$$

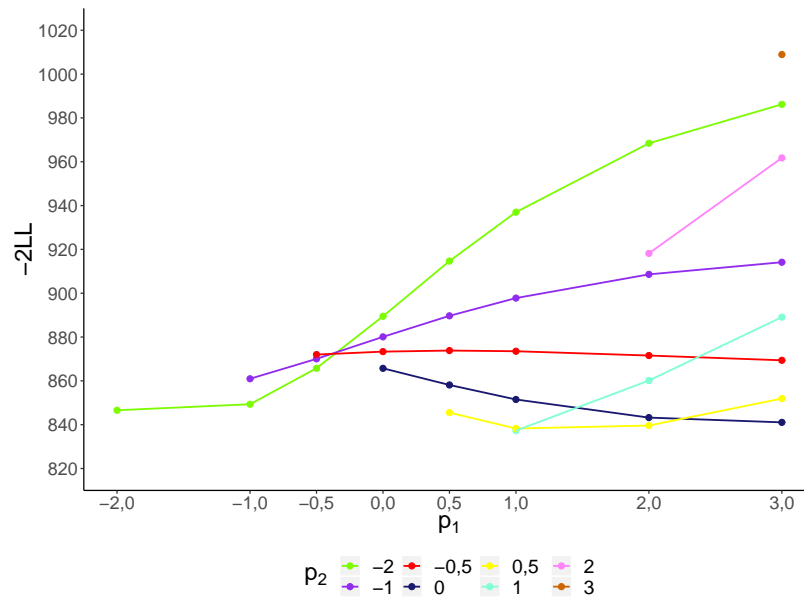


Figura 5.22. Gráfico da componente do desvio estimada pelos modelos (5.11) e (5.12) em função da potência p_1 estratificada por p_2 .

Investigamos uma possível interação entre x_{ij} e w_i inserindo os termos $x_{ij}w_i$ e $x_{ij} \ln(x_{ij})w_i$ no modelo (5.13) e ajustamos o seguinte modelo

$$y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij}. \quad (5.14)$$

Para testar conjuntamente a significância de γ_1 e γ_2 realizamos o teste da razão de verossimilhanças generalizado comparando os modelos encaixados (5.13) e (5.14). Obtivemos $P < 0,001$, indicando que pelo menos uma interação é significativa.

Na Tabela 5.8 encontram-se as estimativas de máxima verossimilhança para os parâmetros do modelo (5.14).

Tabela 5.8. Estimativas de máxima verossimilhança e erros padrão estimados associados aos parâmetros do modelo em (5.14)

Parâmetro	Estimativa	Erro padrão
β_0	-21,90	1,19
β_1	57,36	1,22
β_2	-33,18	0,88
γ_0	-2,07	1,19
γ_1	5,03	1,22
γ_2	1,61	0,88
σ^2	7,16	
σ_0^2	5,58	
<i>AIC</i>	726,29	
<i>BIC</i>	749,82	
Log-verossimilhança	-355,14	

Demos continuidade, portanto, com o modelo (5.14) para a seleção dos efeitos aleatórios.

5.3.3.2 Etapa 2: Seleção dos efeitos aleatórios

Verificamos, utilizando o teste de permutação baseado no BLUP e o teste assintótico baseado na estatística da RVG, quais transformações FP2 do modelo em (5.14)

devem ser incluídas aos perfis individuais. A Tabela 5.9 apresenta os resultados de ambos os testes.

Comparando os modelos (1) e (2) da Tabela 5.9, que testa a inclusão do efeito aleatório b_{1i} no modelo (5.14), o resultado para ambos os testes foi significativo. Estendendo os testes para testar a inclusão do efeito aleatório b_{2i} no modelo que já contém b_{0i} e b_{1i} , tal como os modelos (2) e (3) da Tabela 5.9, mais uma vez, o resultado do teste foi significativo.

Tabela 5.9. Resultados dos testes de permutação e assintótico para inclusão de efeitos aleatórios específicos ao modelar a taxa de ultrafiltração em dialisadores de alto fluxo

Teste	Estatística observada		P	
	BLUP	TRVR	Permutação	Assintótico
(1) vs (2)	31,479	21,816	< 0,0001	< 0,0001
(2) vs (3)	442,201	21,700	< 0,0001	< 0,0001

$$(1): y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + \epsilon_{ij}$$

$$(2): y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + b_{1i} x_{ij} + \epsilon_{ij}$$

$$(3): y_{ij} = \beta_0 + (\beta_1 + \gamma_1 w_i) x_{ij} + (\beta_2 + \gamma_2 w_i) x_{ij} \ln(x_{ij}) + \gamma_0 w_i + b_{0i} + b_{1i} x_{ij} + b_{2i} x_{ij} \ln(x_{ij}) + \epsilon_{ij}.$$

Portanto, temos evidências de que coeficientes relacionados aos termos x_{ij} e $x_{ij} \ln(x_{ij})$ variam entre os dialisadores e devem ser incluídos nos perfis individuais.

A etapa 2 é finalizada com o seguinte modelo

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + \gamma_1 w_i + b_{1i}) x_{ij} + (\beta_2 + \gamma_2 w_i + b_{2i}) x_{ij} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.15)$$

em que

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2),$$

com \mathbf{D} definida em (5.5).

Na Tabela 5.10 encontram-se as estimativas de máxima verossimilhança restrita

para os parâmetros do modelo (5.15).

Tabela 5.10. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo em (5.15)

Parâmetro	Estimativa	Erro padrão
β_0	-22,01	1,26
β_1	57,47	2,06
β_2	-33,26	1,51
γ_0	-1,67	1,81
γ_1	4,59	2,94
γ_2	1,91	2,14
σ^2	3,94	
σ_0^2	4,49	
σ_1^2	27,53	
σ_2^2	14,78	
σ_{01}	-0,99	
σ_{02}	0,94	
σ_{12}	-0,95	
<i>AIC</i>	685,93	
<i>BIC</i>	723,60	
Log-verossimilhança	-329,97	

5.3.3.3 Etapa 3: Seleção de uma estrutura de covariância para os erros

Não foi possível considerar a análise da matriz de covariâncias e correlações amostrais e, conseqüentemente, a do gráfico de linhas dessa matriz, devido as condições de avaliação não serem as mesmas para todas as unidades amostrais. Consideramos então explorar o gráfico dos resíduos condicionais padronizados.

A Figura 5.23 apresenta o gráfico dos resíduos condicionais padronizados, referentes ao modelo (5.15), em função da pressão transmembrana, destacando um padrão de variabilidade crescente nos resíduos com pressões transmembranas mais altas, que já havíamos sinalizado com o gráfico de perfis.

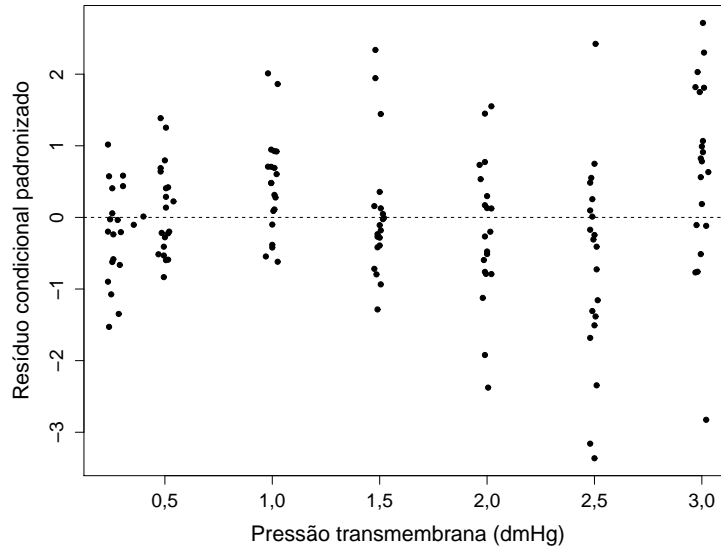


Figura 5.23. Gráfico do resíduo condicional padronizado *versus* pressão transmembrana para o modelo em (5.15).

A fim de capturar essa heterocedasticidade foram testadas algumas funções de variância. A função que forneceu uma representação mais apropriada foi a função potência, a mesma selecionada para o modelo (5.9). Reajustamos, portanto, o seguinte modelo

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + \gamma_1 w_i + b_{1i}) x_{ij} + (\beta_2 + \gamma_2 w_i + b_{2i}) x_{ij} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.16)$$

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim N(0, \sigma^2 x_{ij}^{2\delta}),$$

com \mathbf{D} definida em (5.5).

Para o modelo (5.16) obtivemos um padrão razoavelmente homogêneo de variabilidade para os resíduos condicionais padronizados (Figura 5.24).

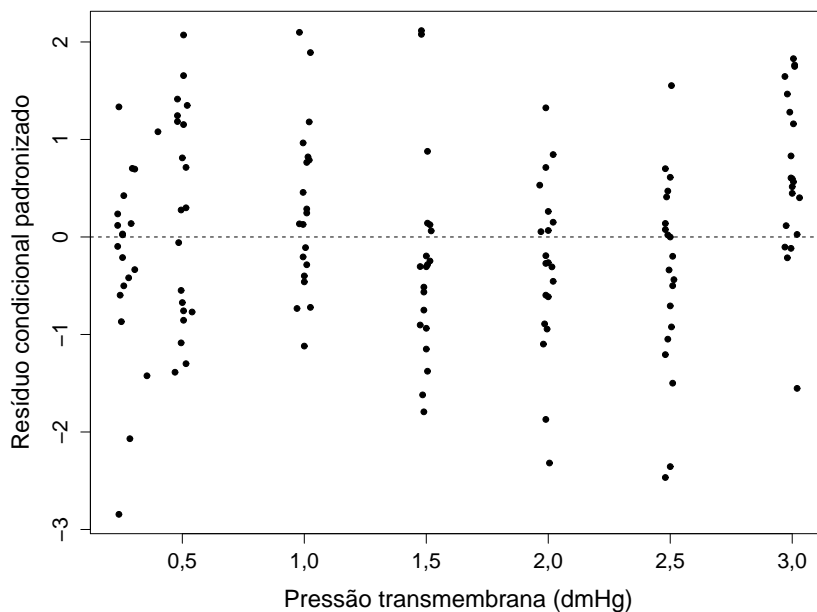


Figura 5.24. Gráfico do resíduo condicional padronizado em função da pressão transmembrana para o modelo em (5.16).

Análises gráficas dos resíduos para o modelo (5.16) são apresentadas nas Figuras 5.25–5.29.

Analisando a Figura 5.25, não se observa nenhum tipo de tendência do resíduo marginal, indicando que não há falta de ajuste da parte que modela $\mathbb{E}(Y)$.

Pela Figura 5.26, não se percebe afastamento importante da normalidade por parte do resíduo com confundimento mínimo, indicando ser plausível a suposição de normalidade do erro condicional.

O gráfico para a estrutura de covariância, Figura 5.27, indica que a estrutura de covariância adotada no modelo (5.16) está adequada.

A Figura 5.28, o gráfico das distâncias de Mahalanobis *versus* os índices dos dialisadores, não indica nenhuma unidade amostral discrepante e o gráfico dos quantis dessas distâncias (Figura 5.29) detecta os dialisadores 3 e 6 com valores muito baixos, o que afetou levemente a suposição de normalidade dos efeitos aleatórios.

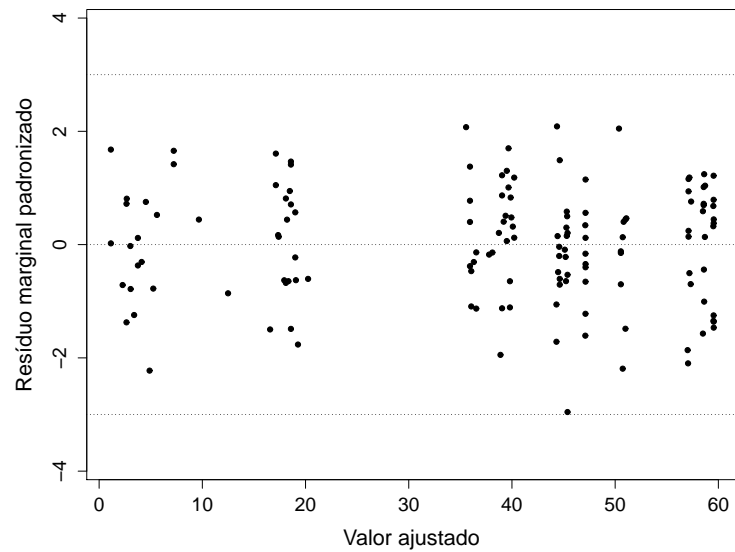


Figura 5.25. Gráfico do resíduo marginal padronizado *versus* valor ajustado para o modelo em (5.16).

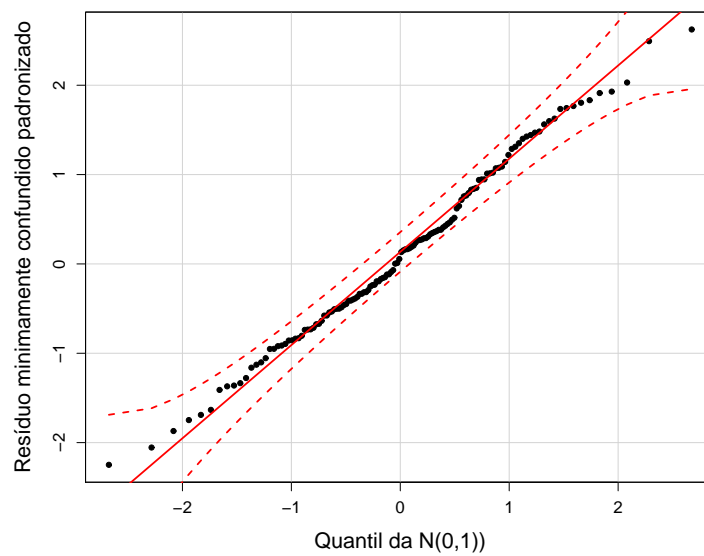


Figura 5.26. Gráfico do envelope simulado com 95% para resíduos condicionais minimamente confundidos para o modelo em (5.16).

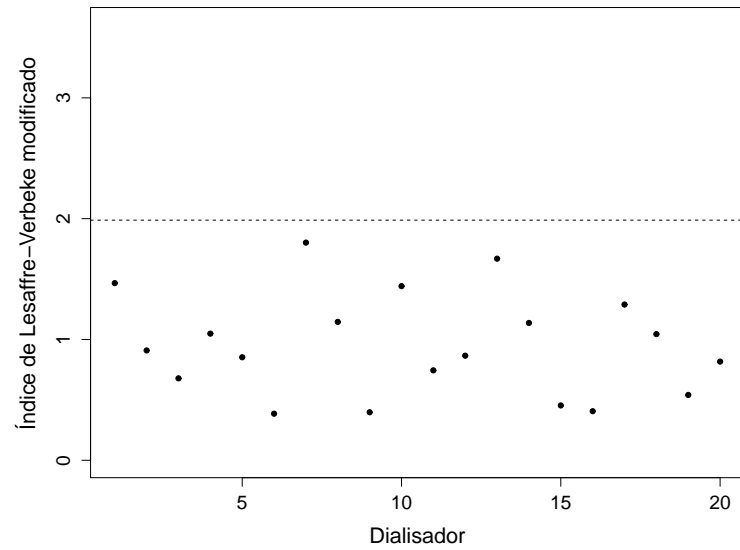


Figura 5.27. Gráfico do índice de Lesaffre-Verbeke modificado *versus* índice do dialisador para o modelo em (5.16).

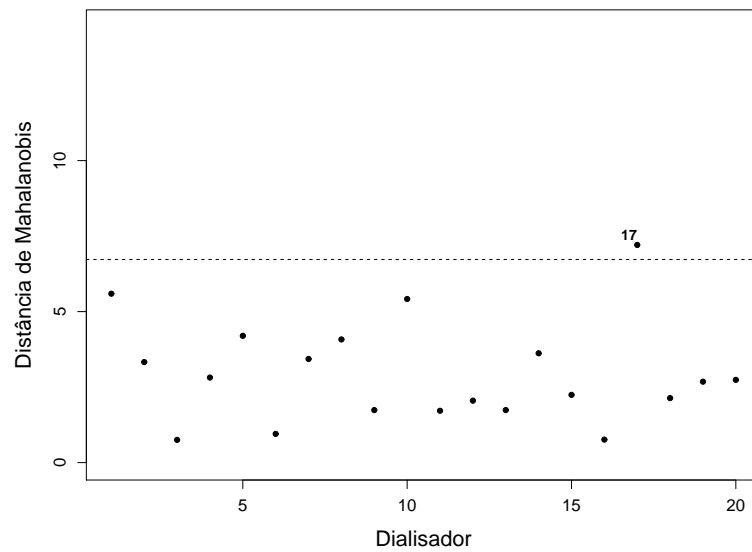


Figura 5.28. Gráfico baseado na distância de Mahalanobis *versus* índice do dialisador para o modelo em (5.16).

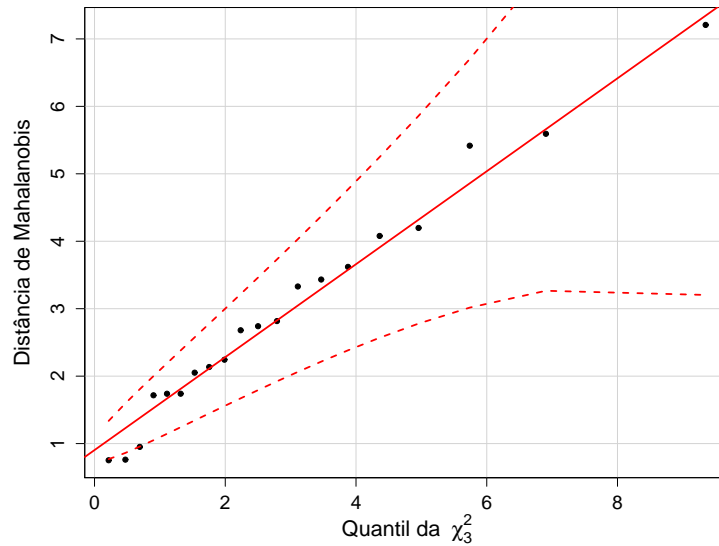


Figura 5.29. Gráfico QQ baseado na distância de Mahalanobis correspondente ao modelo em (5.16).

5.3.3.4 Etapa 4: Redução do modelo

Nesta etapa, verificamos se a transformação FP2 na variável regressora pressão transmembrana selecionada na etapa 1, ou seja, x_{ij} e $x_{ij} \ln(x_{ij})$, sofre alguma alteração. Na etapa 1 foi assumido modelo de independência condicional homocedástico, agora, utilizando o método da máxima verossimilhança, ajustamos o modelo

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + b_{1i}) x_{ij}^{p_1} + (\beta_2 + b_{2i}) x_{ij}^{p_2} + \epsilon_{ij}, \quad (5.17)$$

quando $p_1 \neq p_2$, e

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + b_{1i}) x_{ij}^{p_1} + (\beta_2 + b_{2i}) x_{ij}^{p_1} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.18)$$

quando $p_1 = p_2$, ambos com

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}) \quad \text{e} \quad \epsilon_{ij} \sim N(0, \sigma^2 x_{ij}^{2\delta}), \quad (5.19)$$

com a matriz \mathbf{D} não estruturada.

O melhor modelo PFMM2, selecionado pelo método da máxima verossimilhança, continuou tendo potências (1, 1) para x_{ij} , logo permanecemos com as transformações FP2 x_{ij} e $x_{ij} \ln(x_{ij})$. Seguindo os passos do algoritmo MFPI, definimos as novas variáveis regressoras $x_{ij}w_i$ e $x_{ij} \ln(x_{ij})w_i$, mas para testar a significância dos termos associados às taxas de fluxo de sangue utilizamos o teste F sequencial que leva em conta a ordem em que foram inseridos no modelo. Foi obtido $P = 0,1742$ para a interação $x_{ij} \ln(x_{ij})w_i$, sugerindo que esse termo não contribui significativamente para o ajuste do modelo. Portanto, o modelo PFMM2(1,1), com $AIC = 653,06$, fica dado por

$$y_{ij} = (\beta_0 + \gamma_0 w_i + b_{0i}) + (\beta_1 + \gamma_1 w_i + b_{1i}) x_{ij} + (\beta_2 + b_{2i}) x_{ij} \ln(x_{ij}) + \epsilon_{ij}, \quad (5.20)$$

em que

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}_3(\mathbf{0}, \sigma^2 \mathbf{D}), \quad \epsilon_{ij} \sim N(0, \sigma^2 x_{ij}^{2\delta}),$$

com a matriz \mathbf{D} não estruturada

Na Tabela 5.11 encontram-se as estimativas de máxima verossimilhança restrita para os parâmetros do modelo (5.20).

Tabela 5.11. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros do modelo (5.20)

Parâmetro	Estimativa	Erro padrão
β_0	-23,01	0,79
β_1	58,23	1,35
β_2	-33,89	0,98
γ_0	-2,67	0,79
γ_1	7,03	0,80
σ^2	1,59	
σ_0^2	7,56	
σ_1^2	28,47	
σ_2^2	15,57	
σ_{01}	-0,87	
σ_{02}	0,80	
σ_{12}	-0,95	
δ	0,86	
<i>AIC</i>	650,29	
<i>BIC</i>	688,05	
Log-verossimilhança	-312,14	

Na Figura 5.30 pode-se notar que a resposta média da taxa de fluxo de sangue de 300 dl/min é maior que a taxa de fluxo de sangue de 200 dl/min e que esse aumento parece iniciar, aproximadamente, a partir de pressões maiores que 0,5 dmHg. Ambas as curvas mostram ponto de máximo, que se dá para pressão levemente acima de 2,0 dmHg, quando o fluxo é 200 dl/min, e levemente acima de 2,5 dmHg, quando o fluxo é 300 dl/min.

A Figura 5.31 ilustra graficamente a taxa de variação da curva de crescimento, medida pela primeira derivada do preditor linear do modelo em (5.20), para as taxas de fluxo de sangue de 200 dl/min e 300 dl/min. Notamos que a taxa de variação dos níveis de ultrafiltração, tanto para o fluxo de sangue de 200 dl/min quanto para o de 300 dl/min, vai diminuindo conforme a pressão aumenta, sendo no início um pouco mais rápida e depois caindo vagarosamente. Para ambas as curvas não percebemos

indicação de estabilidade.

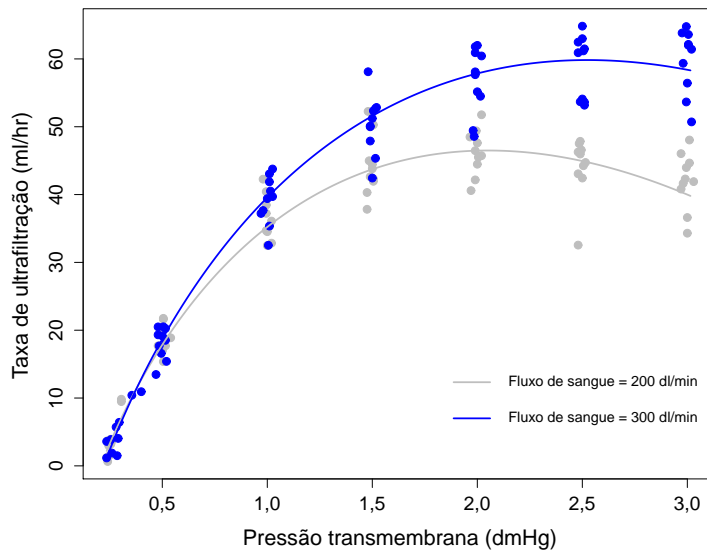


Figura 5.30. Gráfico do perfil médio com curvas ajustadas pelos valores preditos do modelo em (5.20) por taxa de fluxo de sangue de 200 dl/min e de 300 dl/min.

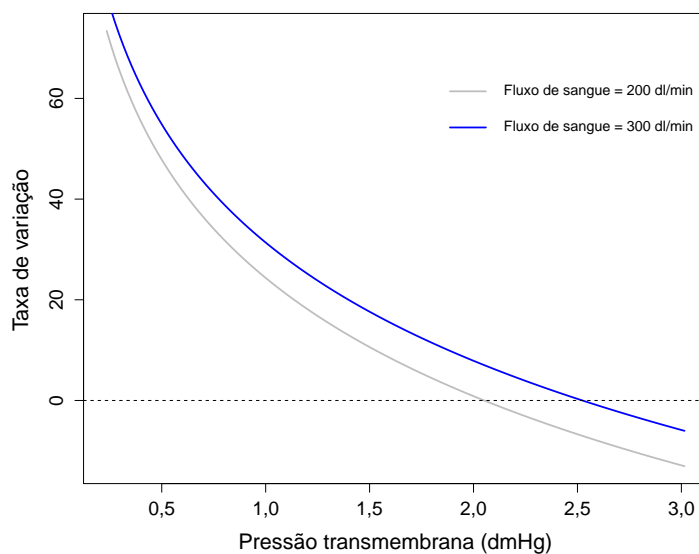


Figura 5.31. Taxa de variação em função da pressão transmembrana para os níveis de fluxo de sangue de 200 dl/min e 300 dl/min.

As curvas individuais ajustadas pelos modelos (5.9) e (5.20) são apresentadas

na Figura 5.32. Em termos de curvas ajustadas, as diferenças são sutis entre estes dois modelos. No entanto, o exemplo ilustra a flexibilidade dos FPs para modelar relações curvas evitando polinômios de alta ordem.

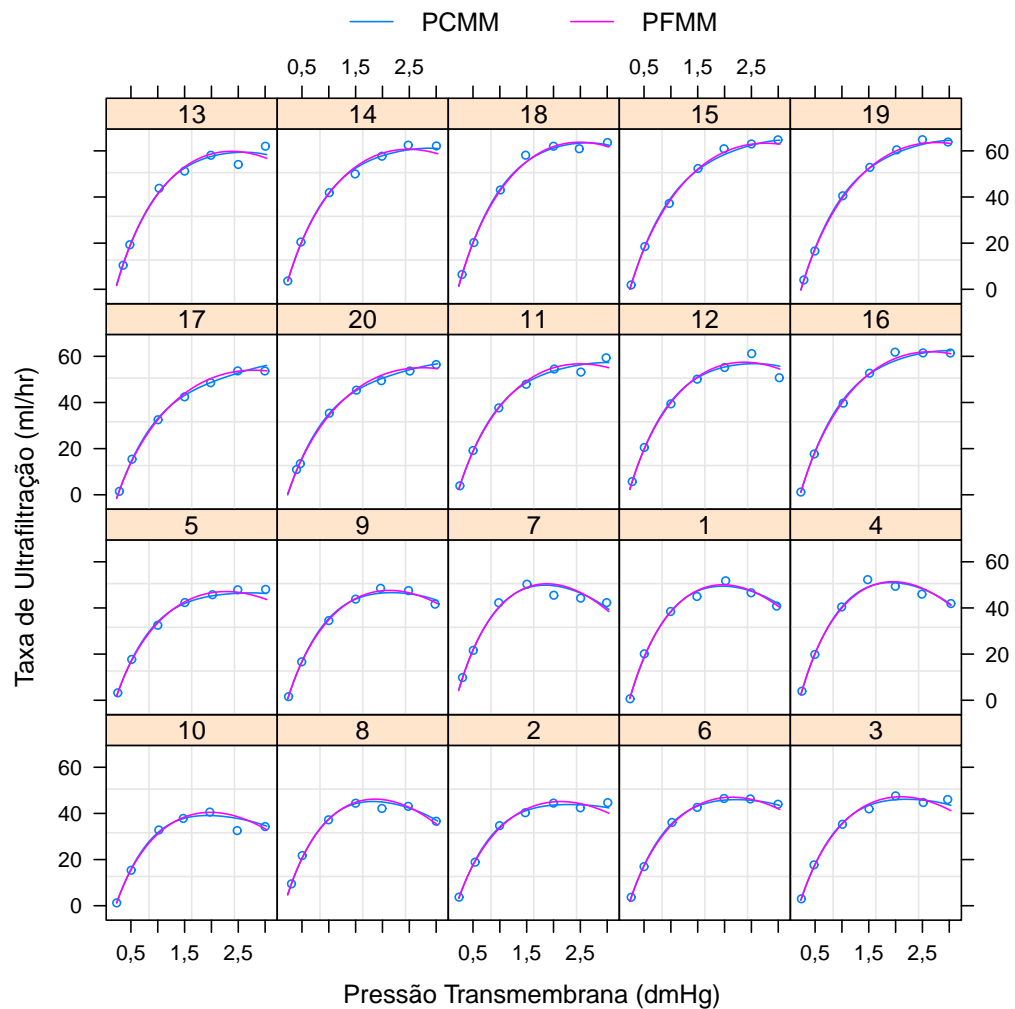


Figura 5.32. Curvas individuais ajustadas da taxa de ultrafiltração em função da pressão transmembrana correspondentes aos modelos (5.9), PCMM, e (5.20), PFMM.

Na Figura 5.33 temos a ilustração de gráficos de diagnóstico do modelo em (5.20). Comparando esses gráficos aos das Figuras 5.25–5.29 observamos poucas alterações entre eles. Gráficos de influência para o modelo em (5.20) são apresentados nas Figuras 6.8–6.14 do Anexo e não indicam observações ou unidades influentes.

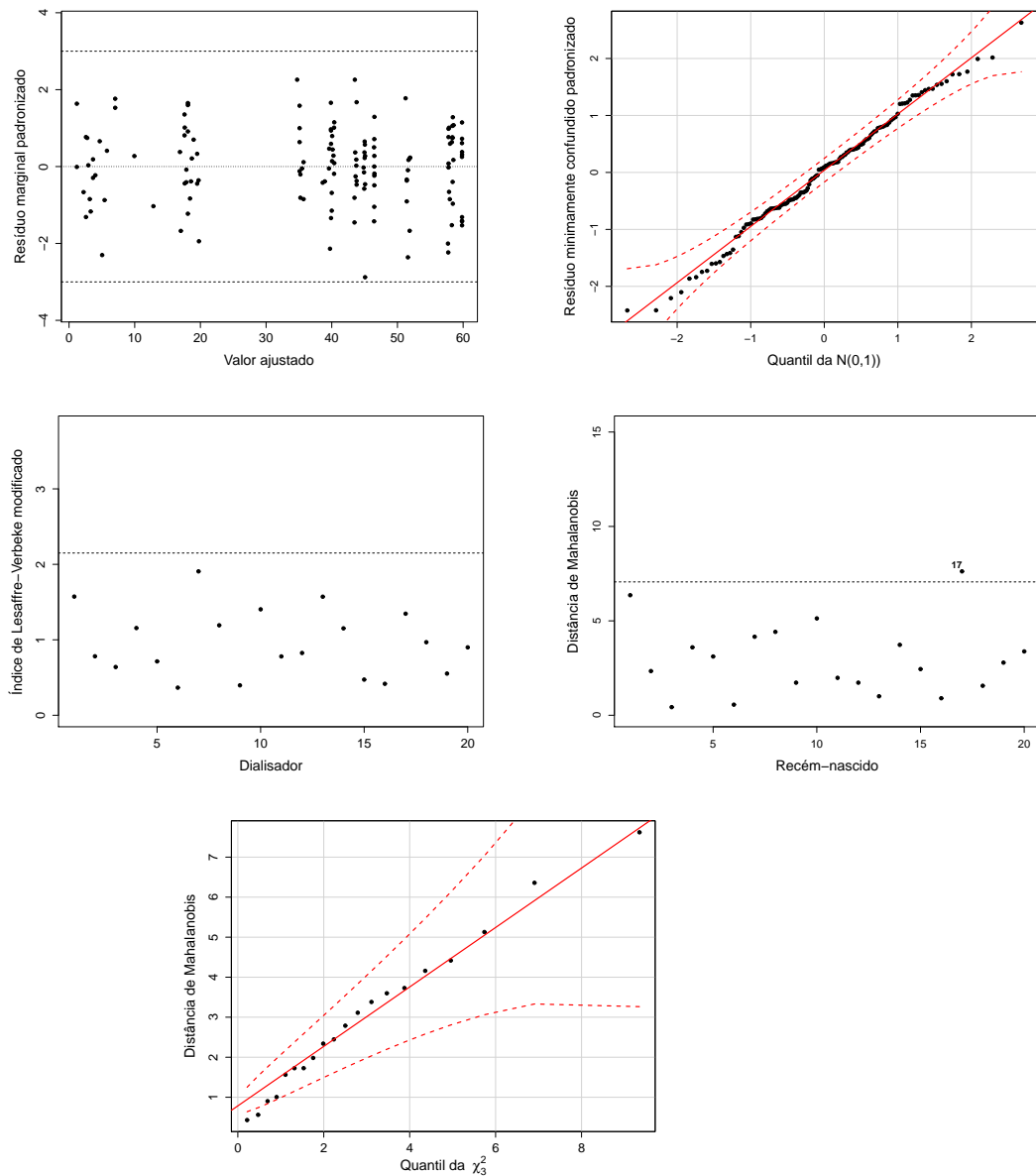


Figura 5.33. Gráficos de diagnóstico para o modelo em (5.20).

Utilizando o critério AIC para comparar os modelos (5.9) e (5.20), observamos que os resultados obtidos são bastante similares e que as diferenças entre as curvas geradas pelos dois modelos são muito pequenas. O modelo misto com polinômios convencionais precisou de um polinômio de quarta ordem para obter uma aproximação da verdadeira função de regressão, sendo atraente, não só por poder ser

ajustado por meio de metodologia amplamente disseminada, mas também porque dispõe de técnicas de diagnóstico bem estabelecidas. O modelo polinomial fracionário de efeitos mistos, por outro lado, além de ter as mesmas vantagens dos polinômios convencionais, mostrou ser uma alternativa bastante parcimoniosa, lembrando que o foco principal dessa tese é extrair informações completas de variáveis contínuas, resultando em modelos com formas funcionais plausíveis.

5.4 Aplicação 3: Efeitos da poluição sobre o preço de residências

5.4.1 Descrição do estudo

Na terceira aplicação analisamos um conjunto de dados originalmente considerado em Harrison & Rubinfeld (1978) que estudaram o impacto da poluição do ar, medida pela concentração de óxido de nitrogênio, sobre o preço de residências ocupadas por seus proprietários. Na análise os autores incluíram ainda outras 13 variáveis regressoras, indicadoras de qualidade e relevantes para explicar a variável resposta preço.

O estudo utilizou setores censitários selecionados aleatoriamente de uma amostra de localidades (*towns*) do estado norte-americano de Massachusetts, que à época, 1970, era designada de Área Estatística Metropolitana Padrão de Boston. Por uma questão de uniformização designamos *towns* por localidades, e não por cidades, visto que o conjunto de dados indica tanto setores censitários amostrados de bairros/distritos da cidade Boston (capital e maior cidade de Massachusetts) quanto de cidades de Massachusetts.

Dentro desse cenário, o conjunto de dados totaliza 506 setores censitários amostrados de 92 localidades, assim distribuídos: 132 setores censitários amostrados de 15 localidades de uma única cidade, Boston, e 374 setores censitários amostrados de 77 localidades, sendo cada localidade referente a uma cidade de Massachusetts.

As 92 localidades estão, portanto, associadas a 78 cidades de Massachusetts

distribuídas em cinco condados, são eles: Middlesex (30 cidades), Norfolk (20 cidades), Essex (15 cidades), Plymouth (9 cidades) e Suffolk (4 cidades). O estado de Massachusetts possui 14 condados e cada condado abriga muitas cidades. Com base na Figura 5.34, nota-se que os cinco condados do estudo são fronteiriços.



Figura 5.34. Mapa do estado de Massachusetts com seus 14 condados.

Os dados exibem uma estrutura hierárquica e nessa tese foram considerados dois níveis de hierarquia com setores censitários agrupados dentro de localidades. O nível 1, sendo a unidade de observação, em que está o setor censitário e, o nível 2, sendo o grupo, em que está a localidade. Esta estrutura implica que os preços das casas nos setores censitários dentro da uma mesma localidade provavelmente são correlacionados.

O conjunto de dados contém 506 observações e 14 variáveis, sendo que nove foram observadas no nível 1 e cinco no nível 2. Os nomes e uma breve descrição das variáveis, associadas a cada nível, são apresentados a seguir. Mais detalhes podem ser encontrados em Harrison & Rubinfeld (1978).

Variáveis no nível 1 (setor censitário):

LMV Logaritmo do valor mediano das casas ocupadas pelos proprietários, em USD1.000.

AGE Proporção de unidades ocupadas pelos proprietários construídas antes de 1940.

BK Proporção de habitantes negros americanos.

CRIM Taxa de criminalidade per capita.

CHAS Variável *dummy* com valor 1 se o setor censitário for limitado pelo Rio Charles e 0 caso contrário.

DIS Distância ponderada para cinco centros de emprego na região de Boston.

LSTAT Proporção da população com baixa renda (proporção de adultos sem educação secundária e proporção de trabalhadores do sexo masculino classificados como operários).

NOX Concentração média anual de óxido de nitrogênio em partes por 10 milhões.

RM Número médio de quartos por habitação.

Variáveis no nível 2 (localidade):

INDUS Proporção de acres de comércio não varejista.

PTRATIO Proporção aluno-professor.

RAD Índice de acessibilidade para rodovias radiais.

TAX Taxa do imposto sobre o valor total da propriedade (mede o custo dos serviços públicos em cada comunidade), em USD10.000.

ZN Proporção de lotes residenciais com tamanho superior a 25.000 pés².

O número de unidades de observação é desbalanceado, variando de 1 até 30 setores censitários por localidade. A variável resposta é o logaritmo do valor mediano das casas ocupadas pelos proprietários no setor censitário e o conjunto de dados possui 12 variáveis regressoras contínuas e uma binária.

Esse conjunto de dados apresenta características ricas que permitem a exploração de diversas metodologias além da modelagem clássica (Gauss-Markov). Por essa

razão, tem sido utilizado para ilustrar diferentes aspectos metodológicos da modelagem estatística. Por exemplo, Belsley et al. (1980) ilustraram os efeitos do uso de estratégias robustas de regressão; Breiman & Friedman (1985) utilizaram estimação não paramétrica, por meio do algoritmo ACE (*Alternating Conditional Expectation*); Lange & Ryan (1989) consideraram um método para verificar a adequação do modelo e, em particular, a suposição distributiva sobre os efeitos aleatórios; Pace (1993) considerou estimadores *kernel* não paramétricos; Zhong et al. (2002) estudaram a combinação dos modelos mistos com covariáveis medidas com erros; Peña (2005) definiu uma nova estatística para medir a influência de uma observação; Zare & Rasekh (2011) derivaram diferentes medidas de diagnóstico para modelos mistos com erros nas covariáveis e Singer et al. (2017), com base na interpretação de gráficos de diagnóstico, procederam à identificação de medidas corretivas para possíveis violações dos pressupostos do modelo, bem como o uso de método de estimação robusta.

No processo de realização de outro estudo, Gilley & Pace (1996) compararam os dados de Harrison & Rubinfeld (1978) aos dados originais do censo e encontraram 8 observações da variável resposta com valor incorreto (Tabela 5.12).

O conjunto de dados completo, com as correções na variável resposta, encontra-se disponível em <http://lib.stat.cmu.edu/datasets/> e também no pacote `mlbench` (Leisch & Dimitriadou, 2010) do software R, com o nome de `BostonHousing2`. Nesta aplicação foi utilizada a versão corrigida dos dados.

Tabela 5.12. Observações da variável dependente que foram codificadas erroneamente

Observação e Setor censitário	Valor mediano	Valor mediano corrigido	% de erro
8 – 2042	27,1	22,1	22,62
39 – 2084	24,7	24,2	2,07
119 – 3585	37,0	33,0	12,12
241 – 3823	22,0	27,0	-18,42
438 – 0905	8,7	8,2	6,10
443 – 0911	18,4	14,8	24,32
455 – 0923	14,9	14,4	3,47
506 – 1805	11,9	19,0	-37,37

Harrison & Rubinfeld (1978), Belsley et al. (1980), Singer et al. (2017) e outros consideram formas funcionais pré-definidas para as variáveis regressoras as quais encontram-se indicadas em Harrison & Rubinfeld (1978). A nossa proposta é, utilizando o algoritmo MFP, encontrar a forma funcional de cada variável regressora partindo dos valores originalmente observados. Embora as características dos dados conduzam-nos diretamente à proposta de um modelo misto, para justificar nosso procedimento, partiremos do modelo clássico, cujo diagnóstico levanta suspeitas e sugere correções na modelagem.

5.4.2 Modelo de regressão múltipla

O modelo utilizado por Harrison & Rubinfeld (1978), apresentado e discutido por Belsley et al. (1980), é um modelo de regressão linear múltipla cuja especificação geral para uma observação individual do logaritmo do valor mediano da casa (LMV_{ij}), relativo ao j -ésimo setor censitário na i -ésima localidade, é dado por

$$\begin{aligned} \text{LMV}_{ij} = & \beta_0 + \beta_1 \text{CRIM}_{ij} + \beta_2 \text{ZN}_{ij} + \beta_3 \text{INDUS}_{ij} + \beta_4 \text{CHAS}_{ij} \\ & + \beta_5 \text{NOX}_{ij} + \beta_6 \text{RM}_{ij} + \beta_7 \text{AGE}_{ij} + \beta_8 \text{DIS}_{ij} + \beta_9 \text{RAD}_{ij} \\ & + \beta_{10} \text{TAX}_{ij} + \beta_{11} \text{PTRATIO}_{ij} + \beta_{12} \text{BK}_{ij} + \beta_{13} \text{LSTAT}_{ij} + \epsilon_{ij}, \end{aligned}$$

em que $i = 1, \dots, 92$, $j = 1, \dots, n_i$ e assumindo $\epsilon_{ij} \sim \mathcal{NN}(0, \sigma^2)$. Para simplificar a notação vamos reescrever o modelo como

$$\begin{aligned} y_{ij} = & \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} \\ & + \beta_6 x_{6ij} + \beta_7 x_{7ij} + \beta_8 x_{8ij} + \beta_9 x_{9ij} + \beta_{10} x_{10ij} \\ & + \beta_{11} x_{11ij} + \beta_{12} x_{12ij} + \beta_{13} x_{13ij} + \epsilon_{ij}. \end{aligned} \tag{5.21}$$

Com essas especificações, podemos escrever o modelo (5.21) na forma matricial por

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}). \tag{5.22}$$

Doravante, utilizaremos a notação matricial dos modelos, especificando onde ocorrerem mudanças.

Para verificar a linearidade dos efeitos fixos no modelo (5.22), consideramos avaliar o gráfico dos resíduos padronizados *versus* valores de cada variável regressora contínua, bem como *versus* valores ajustados. Os gráficos da Figura 5.35 indicam a natureza não linear da relação entre os resíduos padronizados do modelo (5.22) e as variáveis regressoras CRIM, RM, DIS e LSTAT, sugerindo falta de ajuste. Observe na Figura 5.36 que alguns pontos estão fora das bandas de confiança dando indícios de que a suposição de distribuição normal para os erros aleatórios talvez não seja adequada.

Com base na Figura 5.37, que apresenta os boxplots dos resíduos do modelo (5.22) por localidade, vemos que os resíduos não estão centrados em torno do zero e tendem a ser negativos para algumas localidades e positivos para outras. Além disso, percebemos muita variação individual dentro das localidades, com algumas tendo valores de resíduos relativamente mais altos e outras com valores relativamente mais baixos.

Podemos modelar essas diferenças individuais assumindo diferentes interceptos aleatórios para cada localidade.

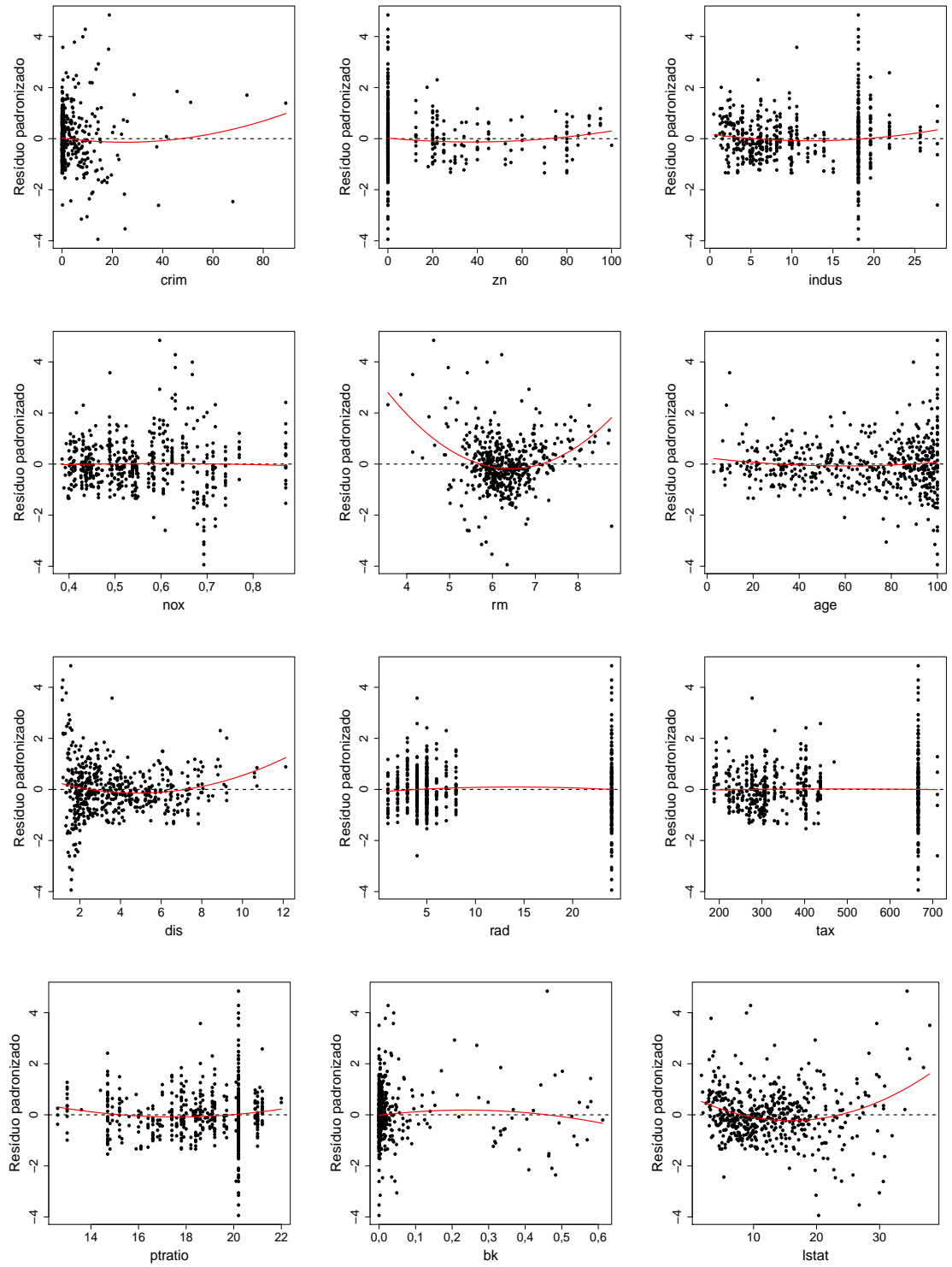


Figura 5.35. Gráficos dos resíduos padronizados *versus* as variáveis regressoras do modelo em (5.22).

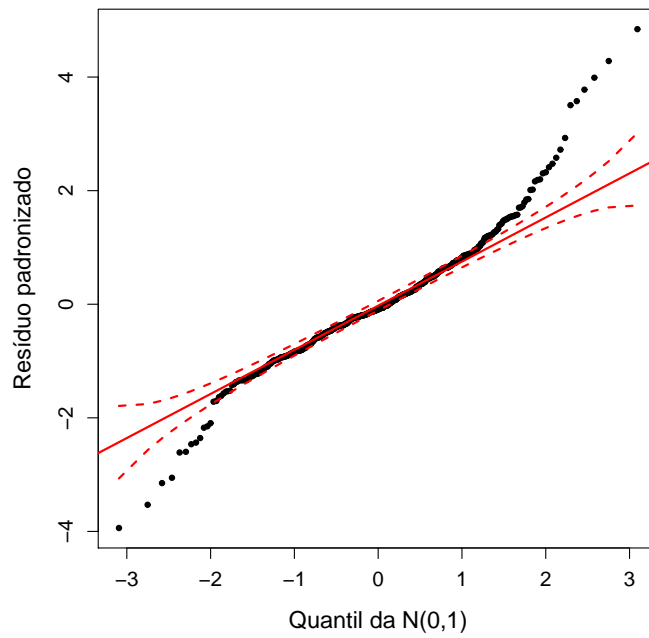


Figura 5.36. Gráfico de probabilidades normal, com envelope gerado, para os resíduos padronizados do modelo em (5.22).

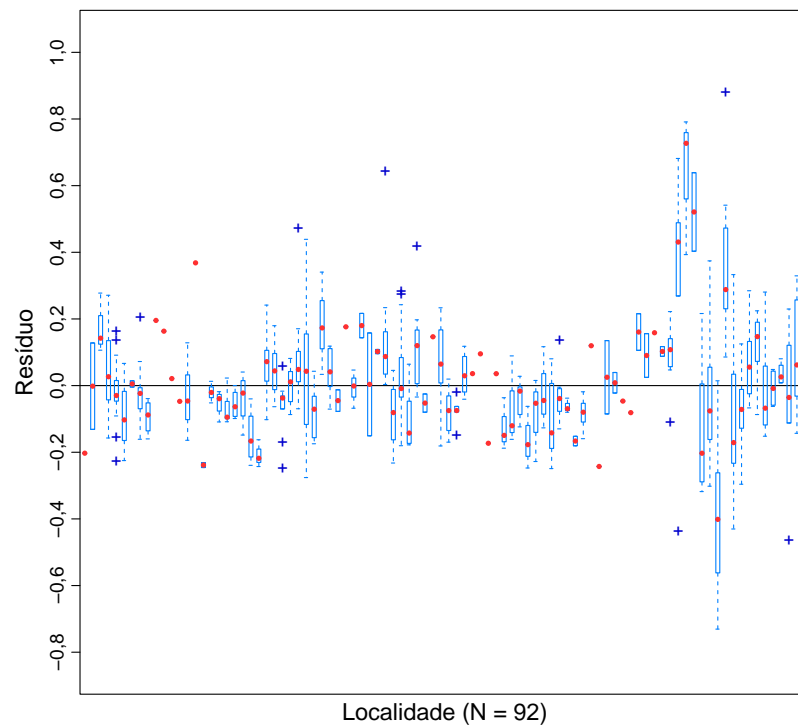


Figura 5.37. Boxplots dos resíduos brutos por localidade correspondente ao modelo em (5.22).

5.4.3 Modelo linear misto

Para tentar melhorar o ajuste do modelo (5.22), foi ajustado o modelo linear misto considerando interceptos aleatórios para as localidades dado por

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_ib_{0i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \quad (5.23)$$

com $b_{0i} \sim \mathcal{N}(0, \sigma_0^2)$ e $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$ denotando, respectivamente, o efeito aleatório da i -ésima localidade e o erro aleatório permanecendo homocedástico.

As estimativas dos parâmetros do modelo (5.23) são apresentadas na Tabela 5.20 e a Figura 5.38 apresenta gráficos de diagnóstico referentes às análises de resíduos para o modelo em (5.23). De forma resumida, podemos dizer que:

1. novamente é possível notar indícios de violação na suposição de linearidade dos efeitos fixos (Figura 5.38(a));
2. a estrutura de covariância proposta parece não ser adequada para 8 localidades, todas do condado de Suffolk, identificadas na Figura 5.38(b) pelos números 76, 77, 78, 80, 81, 82,83 e 90, totalizando 68 setores censitários;
3. nas Figuras 5.38(c) e 5.38(d), os gráficos dos resíduos condicionais padronizados e condicionais minimamente confundidos padronizados, respectivamente, indicam que as suposições de homocedasticidade e normalidade dos erros condicionais não parecem adequadas.

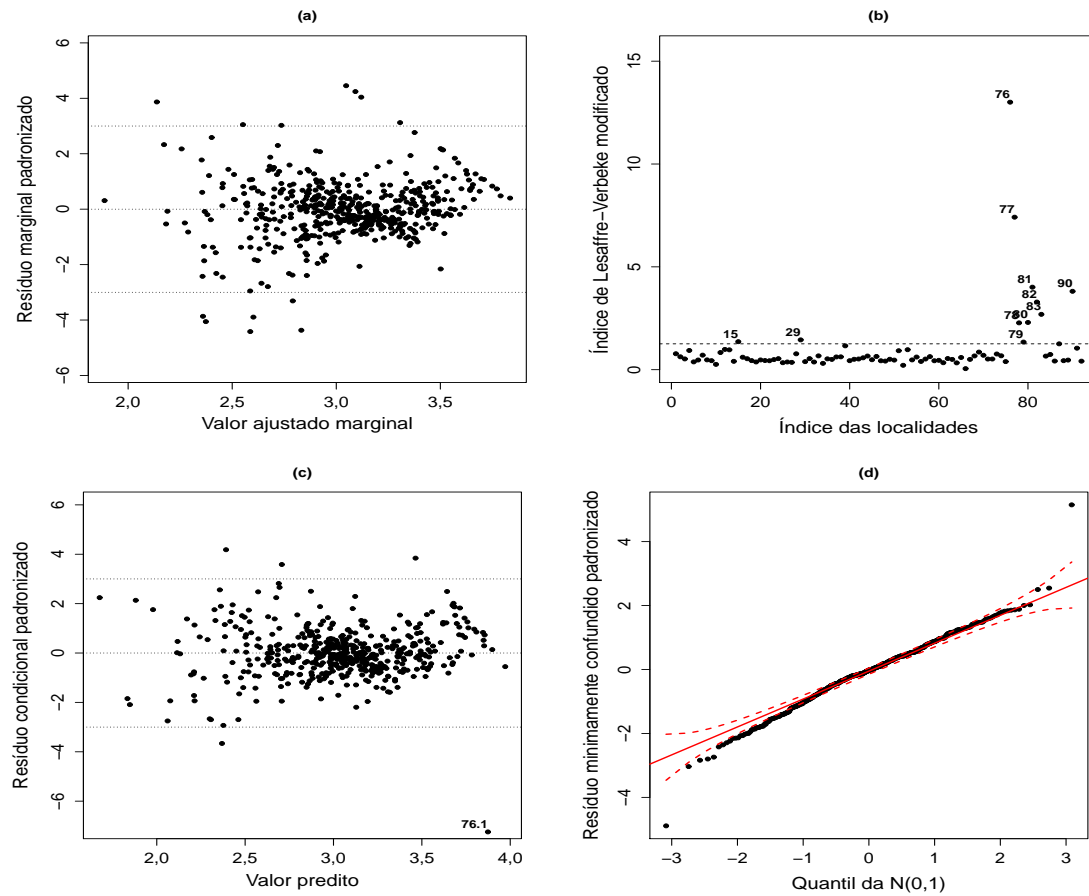


Figura 5.38. Gráficos para análise de resíduos referente ao modelo em (5.23).

5.4.4 Modelo PFMM

De acordo com a análise de resíduos do modelo (5.23), tanto a estrutura da média quanto a estrutura de covariância parecem não ter sido especificadas apropriadamente. A fim de selecionar estruturas de média e de covariância adequadas foi utilizada a estratégia proposta no Capítulo 4.

5.4.4.1 Etapa 1: Seleção de uma estrutura média preliminar

A busca pela estrutura preliminar mais completa possível para a parte fixa do modelo (5.23) foi feita por meio do algoritmo $MFP(\alpha_1, \alpha_2)$. Esse procedimento irá realizar a seleção das variáveis regressoras e determinar sua forma funcional. As

configurações dadas ao algoritmo foram $\alpha_1 = \alpha_2 = 0,05$, com 4 graus de liberdade para cada variável regressora contínua.

Dado que uma transformação FP, x^p , requer $x > 0$, e como as variáveis regressoras BK e ZN possuem valores iguais a zero, utilizamos a transformação $\omega_\delta(x)$, apresentada em (2.7), nessas variáveis regressoras e, daqui em diante, utilizaremos as variáveis regressoras OMEGABK e OMEGAZN que se referem a BK e ZN na escala transformada.

Ajustamos o modelo (5.23) e, como o interesse está na comparação de modelos com diferentes efeitos fixos, a estimação dos parâmetros foi feita pelo método da máxima verossimilhança. Os resultados do ajuste são apresentados na Tabela 5.13. Pelo valor-p apresentado na Tabela 5.13, ou mais precisamente pelo valor absoluto da estatística t , determinamos a ordem de entrada das variáveis regressoras em todos os ciclos do procedimento. Podemos observar que a variável regressora LSTAT é a que apresenta maior efeito e, portanto, será considerada em primeiro lugar e a variável regressora CHAS será a última a ser visitada.

Tabela 5.13. Estimativas de máxima verossimilhança, erro padrão e valor-p ordenados para os parâmetros de regressão do modelo (5.23)

Efeito	Estimativa	Erro padrão	Estatística t	P
LSTAT	-0,020	0,002	-11,749	< 0,001
RM	0,142	0,014	10,332	< 0,001
OMEGABK	-0,367	0,058	-6,276	< 0,001
CRIM	-0,006	0,001	-5,789	< 0,001
NOX	-0,864	0,164	-5,263	< 0,001
RAD	0,020	0,005	4,113	< 0,001
DIS	-0,040	0,011	-3,580	< 0,001
AGE	-0,001	0,000	-3,167	0,002
TAX	-0,001	0,000	-2,989	0,004
PTRATIO	-0,029	0,011	-2,671	0,009
OMEGAZN	0,114	0,098	1,158	0,250
INDUS	0,004	0,005	0,735	0,464
CHAS	-0,018	0,028	-0,622	0,534

Foram necessários apenas 3 ciclos do algoritmo MFP para convergência. O ciclo 1 é apresentado na Tabela 5.14.

Aplicando o passo 1 do FSP para LSTAT, com nível $\alpha_1 = 0,05$, a comparação do modelo FPMM2(-2; 0,5) com o modelo nulo mostra que LSTAT é significativa ($P < 0,001$). Logo, seguimos aplicando os passos 2 e 3 do FSP, com nível $\alpha_2 = 0,05$, para escolher uma função linear ou um FP. O resultado do teste entre o modelo FPMM2(-2; 0,5) e o modelo linear misto também é significativo ($P < 0,001$), enquanto FPMM2(-2; 0,5) *versus* FPMM1(0,5) é não significativo ($P = 0,119$). Portanto, uma função FP1(0,5) é selecionada para LSTAT. Assim, a variável regressora transformada $LSTAT^{0,5}$ é incluída em todos os modelos subsequentes do ciclo 1.

Para a próxima variável regressora, RM, os passos do FSP mostram que uma função FP2 é necessária em um modelo com $LSTAT^{0,5}$ e demais variáveis regressoras, uma vez que (1) FPMM2(0; 0) *versus* nulo é significativo ($P < 0,0001$), (2) FPMM2(0; 0) *versus* modelo linear misto é significativo ($P < 0,001$) e (3) FPMM2(0; 0) *versus* FPMM1(3) também é significativo ($P = 0,023$). Desta forma, uma função FP2(0; 0) é selecionada para RM e, portanto, as variáveis regressoras $\ln(RM)$ e $\ln(RM)\ln(RM)=\ln^2(RM)$ serão incluídas em todos os modelos subsequentes do ciclo 1.

Ajustamos o próximo modelo com $LSTAT^{0,5}$, $\ln(RM)$, $\ln^2(RM)$ e demais variáveis regressoras, OMEGABK é selecionada com uma função linear ($P < 0,0001$). A variável regressora CRIM foi analisada por meio de um modelo com as variáveis regressoras $LSTAT^{0,5}$, $\ln(RM)$, $\ln^2(RM)$, OMEGABK e demais. A comparação do modelo FPMM2(1; 2) com o modelo nulo é significativa ($P < 0,0001$). O teste de FPMM2(1; 2) *versus* modelo linear misto também é significativo ($P = 0,034$), mas FPMM2(1; 2) *versus* FPMM1(0,5) é não significativo ($P = 0,079$). Portanto, uma função FP1(0,5) é selecionada para CRIM. Logo, a variável regressora transformada $CRIM^{0,5}$ é incluída em todos os modelos subsequentes do ciclo 1.

As variáveis regressoras NOX, RAD e DIS entram no modelo com uma função linear. A variável regressora AGE é não significativa com nível de 5% e será excluída

de todos os modelos subsequentes do ciclo 1. As variáveis regressoras TAX e PTRATIO são significativas e serão incluídas no modelo por meio de uma função linear. E, finalmente, as variáveis regressoras OMEGAZN, INDUS e CHAS, com nível de 5% de significância, serão excluídas do modelo. Portanto, ao final do ciclo 1, o modelo compreende as seguintes variáveis regressoras: LSTAT^{0,5}, ln(RM), ln(RM)ln(RM), OMEGABK, CRIM^{0,5}, NOX, RAD, DIS, TAX e PTRATIO.

A Tabela 5.15 apresenta o ciclo 2 onde cada variável regressora é considerada novamente para inclusão ou exclusão. O ciclo 2 inicia com o modelo resultante do ciclo 1. A ordem de visitação permanece a mesma. Podemos notar que todas as variáveis regressoras selecionadas e suas formas funcionais são as mesmas do ciclo 1, mudando somente a transformação FP da variável regressora RM, passando de função FP2(0; 0) para FP2(0; 0,5).

No ciclo 3 (Tabela 5.16), não ocorre nenhuma mudança adicional nas variáveis regressoras ou funções selecionadas. O procedimento finaliza com $-2LL = -520,111$ com as seguintes variáveis regressoras: CRIM^{0,5}, NOX, ln(RM), RM^{0,5}, DIS, RAD, TAX, PTRATIO, OMEGABK, LSTAT^{0,5}. Assim, o modelo PFMM2 fica expresso por

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_{0i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \\ b_{0i} &\sim \mathcal{N}(0, \sigma_0^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}), \end{aligned} \quad (5.24)$$

em que

$$\begin{aligned} \mathbf{Y}_i &= (y_{i1}, y_{i2}, \dots, y_{i,n_i})^\top, \\ \mathbf{X}_i &= \begin{bmatrix} 1 & \text{CRIM}_{i1}^{0,5} & \text{NOX}_{i1} & \ln(\text{RM}_{i1}) & \dots & \text{LSTAT}_{i1}^{0,5} \\ 1 & \text{CRIM}_{i2}^{0,5} & \text{NOX}_{i2} & \ln(\text{RM}_{i2}) & \dots & \text{LSTAT}_{i2}^{0,5} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \text{CRIM}_{i,n_i}^{0,5} & \text{NOX}_{i,n_i} & \ln(\text{RM}_{i,n_i}) & \dots & \text{LSTAT}_{i,n_i}^{0,5} \end{bmatrix}, \\ \boldsymbol{\beta} &= (\beta_0, \beta_1, \dots, \beta_{10})^\top, \\ \mathbf{Z}_i &= \mathbf{1}_{n_i} = (1, 1, \dots, 1)^\top, \\ \boldsymbol{\epsilon}_i &= (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{i,n_i})^\top. \end{aligned}$$

O modelo (5.24) inclui efeitos fixos associados às variáveis regressoras selecionadas pelo procedimento MFP, intercepto aleatório para cada localidade e os erros são assumidos independentes e com mesma variância para todas as localidades. As estimativas para os parâmetros do modelo (5.24) também são apresentadas na Tabela 5.20. A próxima etapa consiste na modelagem dos efeitos aleatórios.

Tabela 5.14. Ciclo 1 do algoritmo MFP para seleção de uma estrutura média preliminar

Variável regressora	Modelo	(vs.)	-2LL	Estatística de teste	P	Potência	(vs.)
LSTAT	Nulo	PFMM2	-356,7395	148,7732	<0,0001*	.	-2; 0,5
	MLM		-480,9567	24,5460	<0,0001 ⁺	1	
	PFMM1		-501,2470	4,2557	0,1191	0,5	
	Final		-501,2470			0,5	
RM	Nulo	PFMM2	-436,8316	85,0141	<0,0001*	.	0; 0
	MLM		-501,2470	20,5987	<0,0001 ⁺	1	
	PFMM1		-514,3316	7,5141	0,0234 ⁺	3	
	Final		-521,8457			0; 0	
OMEGABK	Nulo	PFMM2	-487,5306	36,3198	<0,0001*	.	3; 3
	MLM		-521,8457	2,0047	0,5714	1	
	Final		-521,8457			1	
CRIM	Nulo	PFMM2	-473,1861	57,3148	<0,0001*	.	1; 2
	MLM		-521,8457	8,6552	0,0342 ⁺	1	
	PFMM1		-525,4184	5,0825	0,0788	0,5	
	Final		-525,4184			0,5	
NOX	Nulo	PFMM2	-501,2495	27,2055	<0,0001*	.	-2; -1
	MLM		-525,4184	3,0366	0,3860	1	
	Final		-525,4184			1	
RAD	Nulo	PFMM2	-500,5242	26,0265	<0,0001*	.	-2; -2
	MLM		-525,4184	1,1323	0,7693	1	
	Final		-525,4184			1	
DIS	Nulo	PFMM2	-514,6538	15,6143	0,0036*	.	-2; 0,5
	MLM		-525,4184	4,8497	0,1831	1	
	Final		-525,4184			1	
AGE	Nulo	PFMM2	-521,4432	5,2880	0,2590	.	-2; -2
	Final		-521,4432			.	
TAX	Nulo	PFMM2	-511,8235	12,0404	0,0171*	.	-2; 3
	MLM		-521,4432	2,4207	0,4898	1	
	Final		-521,4432			1	
PTRATIO	Nulo	PFMM2	-513,2508	10,4677	0,0332*	.	-1; -1
	MLM		-521,4432	2,2753	0,5173	1	
	Final		-521,4432			1	
OMEGAZN	Nulo	PFMM2	-520,7816	2,6380	0,6201	.	3; 3
	Final		-520,7816			.	
INDUS	Nulo	PFMM2	-520,0633	4,7339	0,3157	.	0; 0,5
	Final		-520,0633			.	
CHAS	Nulo	MLM	-520,0246	0,0387	0,8440	.	1
	Final		-520,0246			.	

* Teste para a inclusão de uma variável (binária, categórica ou contínua) é significativo com nível α_1

⁺ Teste PFMM2 vs PFMM1 ou PFMM2 vs MLM para a inclusão de uma variável contínua é significativo com nível α_2 .

Tabela 5.15. Ciclo 2 do algoritmo MFP para seleção de uma estrutura média preliminar

Variável regressora	Modelo	(vs.)	-2LL	Estatística de teste	P	Potência	(vs.)
LSTAT	Nulo	PFMM2	-334,0423	188,2666	<0,0001*	.	-0,5; 1
	MLM		-503,7476	18,5613	0,0003+	1	
	FP1		-520,0246	2,2843	0,3191	0,5	
	Final		-520,0246			0,5	
RM	Nulo	PFMM2	-436,3645	83,7467	<0,0001*	.	0; 0,5
	MLM		-495,6113	24,4999	<0,0001+	1	
	PFMM1		-510,0620	10,0492	0,0066+	3	
	Final		-520,1112			0; 0,5	
OMEGABK	Nulo	PFMM2	-490,5658	31,0757	<0,0001*	.	3; 3
	MLM		-520,1112	1,5303	0,6753	1	
	Final		-520,1112			1	
CRIM	Nulo	PFMM2	-466,9703	58,2458	<0,0001*	.	1; 2
	MLM		-516,6608	8,5553	0,0358+	1	
	PFMM1		-520,1112	5,1049	0,0779	0,5	
	Final		-520,1112			0,5	
NOX	Nulo	PFMM2	-496,7054	24,3970	<0,0001*	.	3; 3
	MLM		-520,1112	0,9912	0,8034	1	
	Final		-520,1112			1	
RAD	Nulo	PFMM2	-495,0894	26,4889	<0,0001*	.	-2; 2
	MLM		-520,1112	1,4671	0,6899	1	
	Final		-520,1112			1	
DIS	Nulo	PFMM2	-505,5873	19,8447	0,0005*	.	-2; 1
	MLM		-520,1112	5,3208	0,1498	1	
	Final		-520,1112			1	
AGE	Nulo	PFMM2	-520,1112	5,0288	0,2843	.	-2; -2
	Final		-520,1112			.	
TAX	Nulo	PFMM2	-511,6442	11,4139	0,0223*	.	-2; 3
	MLM		-520,1112	2,9469	0,3999	1	
	Final		-520,1112			1	
PTRATIO	Nulo	PFMM2	-510,2694	12,4126	0,0145*	.	-0,5; -0,5
	MLM		-520,1112	2,5708	0,4626	1	
	Final		-520,1112			1	
OMEGAZN	Nulo	PFMM2	-520,1112	2,6829	0,6122	.	3; 3
	Final		-520,1112			.	
INDUS	Nulo	PFMM2	-520,1112	4,7605	0,3128	.	0; 0,5
	Final		-520,1112			.	
CHAS	Nulo	MLM	-520,1112	0,0425	0,8365	.	1
	Final		-520,1112			.	

* Teste para a inclusão de uma variável (binária, categórica ou contínua) é significativo com nível α_1

+ Teste PFMM2 vs PFMM1 ou PFMM2 vs MLM para a inclusão de uma variável contínua é significativo com nível α_2 .

Tabela 5.16. Ciclo 3 do algoritmo MFP para seleção de uma estrutura média preliminar

Variável regressora	Modelo	(vs.)	-2LL	Estatística de teste	P	Potência	(vs.)
LSTAT	Nulo	PFMM2	-332,2209	190,0783	<0,0001*	.	-0,5; 1
	MLM		-503,9932	18,3060	0,0004 ⁺	1	
	PFMM1		-520,1112	2,1880	0,3349	0,5	
	Final		-520,1112			0,5	
RM	Nulo	PFMM2	-436,3645	83,7467	<0,0001*	.	0; 0,5
	MLM		-495,6113	24,4999	<0,0001 ⁺	1	
	PFMM1		-510,0620	10,0492	0,0066 ⁺	3	
	Final		-520,1112			0; 0,5	
OMEGABK	Nulo	PFMM2	-490,5658	31,0757	<0,0001*	.	3; 3
	MLM		-520,1112	1,5303	0,6753	1	
	Final		-520,1112			1	
CRIM	Nulo	PFMM2	-466,9703	58,2458	<0,0001*	.	1; 2
	MLM		-516,6608	8,5553	0,0358 ⁺	1	
	PFMM1		-520,1112	5,1049	0,0779	0,5	
	Final		-520,1112			0,5	
NOX	Nulo	PFMM2	-496,7054	24,3970	<0,0001*	.	3; 3
	MLM		-520,1112	0,9912	0,8034	1	
	Final		-520,1112			1	
RAD	Nulo	PFMM2	-495,0894	26,4889	<0,0001*	.	-2; 2
	MLM		-520,1112	1,4671	0,6899	1	
	Final		-520,1112			1	
DIS	Nulo	PFMM2	-505,5873	19,8447	0,0005*	.	-2; 1
	MLM		-520,1112	5,3208	0,1498	1	
	Final		-520,1112			1	
AGE	Nulo	PFMM2	-520,1112	5,0288	0,2843	.	-2; -2
	Final		-520,1112			.	
TAX	Nulo	PFMM2	-511,6442	11,4139	0,0223*	.	-2; 3
	MLM		-520,1112	2,9469	0,3999	1	
	Final		-520,1112			1	
PTRATIO	Nulo	PFMM2	-510,2694	12,4126	0,0145*	.	-0,5; -0,5
	MLM		-520,1112	2,5708	0,4626	1	
	Final		-520,1112			1	
OMEGAZN	Nulo	PFMM2	-520,1112	2,6829	0,6122	.	3; 3
	Final		-520,1112			.	
INDUS	Nulo	PFMM2	-520,1112	4,7605	0,3128	.	0; 0,5
	Final		-520,1112			.	
CHAS	Nulo	MLM	-520,1112	0,0425	0,8365	.	1
	Final		-520,1112			.	

* Teste para a inclusão de uma variável (binária, categórica ou contínua) é significativo com nível α_1

+ Teste PFMM2 vs PFMM1 ou PFMM2 vs MLM para a inclusão de uma variável contínua é significativo com nível α_2 .

5.4.4.2 Etapa 2: Seleção dos efeitos aleatórios

Nesta etapa, testamos se os interceptos aleatórios para as localidades devem ser omitidos do modelo (5.24). Isto será feito ajustando o modelo

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \\ \boldsymbol{\epsilon}_i &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}), \end{aligned} \tag{5.25}$$

que é um caso particular do modelo (5.24). O ajuste desses dois modelos foi realizado pelo método de máxima verossimilhança restrita.

Na comparação do modelo (5.24) com o modelo (5.25) testa-se a inclusão do intercepto como efeito aleatório. A hipótese nula e a hipótese alternativa envolvidas nessa comparação são definidas como

$$H_0 : \sigma_0^2 = 0 \quad \textit{versus} \quad H_a : \sigma_0^2 > 0. \tag{5.26}$$

Ambos os testes, assintótico baseado na estatística da razão de verossimilhanças e de permutação baseado nos BLUPs, apresentaram resultado significativo, ($P < 0,001$), o que implica benefício ao manter o intercepto como efeito aleatório para localidade nos modelos subsequentes. Os efeitos aleatórios no intercepto representam a heterogeneidade natural entre as localidades decorrente de fatores não mensurados.

5.4.4.3 Etapa 3: Seleção de uma estrutura de covariância para os erros

Uma vez que os efeitos fixos e aleatórios foram selecionados, no passo 3 selecionamos a estrutura de covariância para os erros do modelo (5.24). Tomando por base o estudo de Singer et al. (2017), que consideraram uma análise residual detalhada para o modelo misto dos dados das habitações de Boston, foram construídos gráficos de resíduos para o modelo (5.24), apresentados nas Figuras 5.39, 5.40 e 5.41.

A Figura 5.39 destaca praticamente as mesmas localidades já indicadas na

Figura 5.38(b) e sugere a não adequação da estrutura de covariância adotada, ou seja, $\mathcal{R}_i = \sigma^2 \mathbf{I}_{n_i}$. Os gráficos dos resíduos condicionais padronizados (Figura 5.40) e minimamente confundidos padronizados (Figura 5.41) sugerem que há indícios de violação nas suposições de homocedasticidade e normalidade dos erros condicionais, respectivamente.

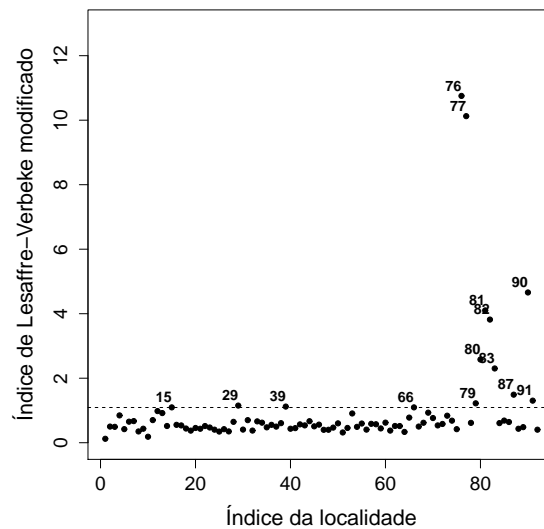


Figura 5.39. Gráfico do índice de Lesaffre-Verbeke modificado *versus* os índices das localidades, modelo (5.24).

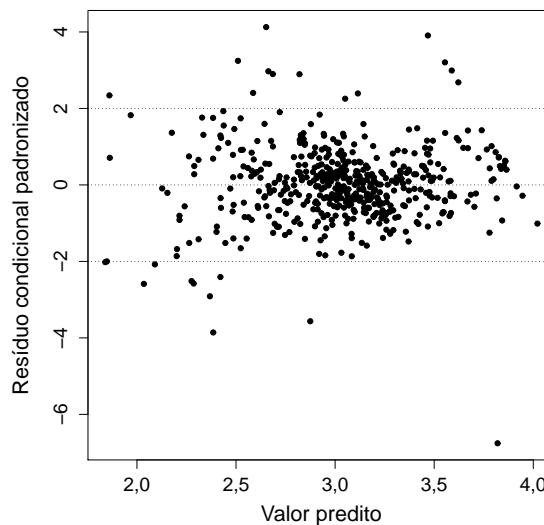


Figura 5.40. Gráfico dos resíduos condicionais padronizados *versus* os valores preditos, modelo (5.24).

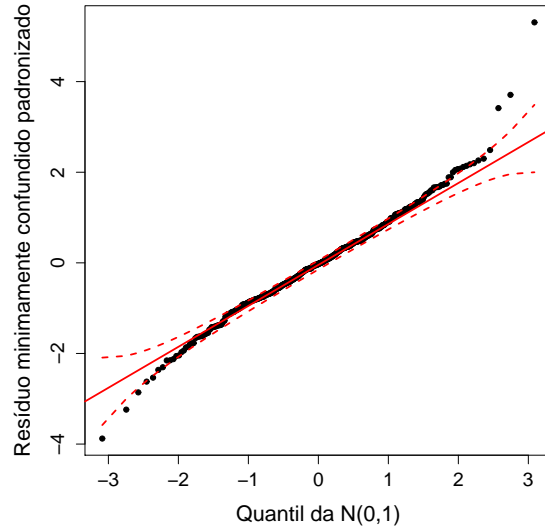


Figura 5.41. Gráfico do resíduo minimamente confundido *versus* quantis da distribuição $N(0,1)$, modelo (5.24).

Como em Singer et al. (2017), incorporamos uma estrutura de correlação de simetria composta intra-unidades amostrais e uma função de variância que atribui variâncias diferentes para cada uma das 7 localidades em destaque na Figura 5.39, que são as localidades 76, 77, 80, 81, 82, 83 e 90, e uma variância comum para cada uma das 85 localidades restantes.

Temos, portanto, um novo agrupamento no conjunto de dados que são localidades dentro de grupos. Assim, para os diferentes grupos ajustamos o seguinte modelo PFMM,

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i b_{0i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \\ b_{0i} &\sim \mathcal{N}(0, \sigma_0^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \delta_{si}^2 \mathbf{R}^*), \quad s = 1, \dots, 8, \end{aligned} \quad (5.27)$$

em que s é a variável de estratificação, nesse caso, grupo; δ_{si} é a função de variância que atribuirá variâncias diferentes para cada grupo, tal que para $s > 1$, $\delta_{si} > 0$, com $i = 76, 77, 80, 81, 82, 83, 90$, e $\delta_{1i} = 1$, caso contrário; e \mathbf{R}^* é a matriz de dimensão $(n_i \times n_i)$ representando a estrutura de correlação, de simetria composta, com um

único parâmetro, ρ , representando a correlação constante nos logaritmos dos preços das residências nos diferentes setores censitários para a mesma localidade.

No entanto, o gráfico do índice de Lesaffre-Verbeke modificado para o modelo (5.27), e outros que foram ajustados sequencialmente incorporando outras localidades, sugeriram a não adequação da estrutura de covariância para um total de mais 16 localidades, além das oito já incorporadas, que foram: 4, 15, 28, 29, 31, 36, 39, 41, 58, 66, 79, 84, 85, 86, 87 e 91.

Atribuindo diferentes variâncias para essas localidades adicionais, temos o seguinte modelo

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i b_{0i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \\ b_{0i} &\sim \mathcal{N}(0, \sigma_0^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \delta_{si}^2 \mathbf{R}^*), \quad s = 1, \dots, 24, \end{aligned} \tag{5.28}$$

tal que para $s > 1$, $\delta_{si} > 0$, com $i = 4, 15, 28, \dots, 87, 90, 91$, e $\delta_{1i} = 1$, em caso contrário; e \mathbf{R}^* definida em (5.27). As estimativas para os parâmetros do modelo (5.28) também são apresentadas na Tabela 5.20.

A Figura 5.42 apresenta o gráfico do índice de Lesaffre-Verbeke modificado para o modelo em (5.28), mostrando um padrão aceitável para a estrutura de covariância proposta. O gráfico do resíduo condicional minimamente confundido padronizado, apresentado na Figura 5.43, mostra padrão coerente com a suposição de normalidade para os erros condicionais. Além disso, o gráfico das distâncias de Mahalanobis (Figura 5.44) indica a normalidade dos efeitos aleatórios.

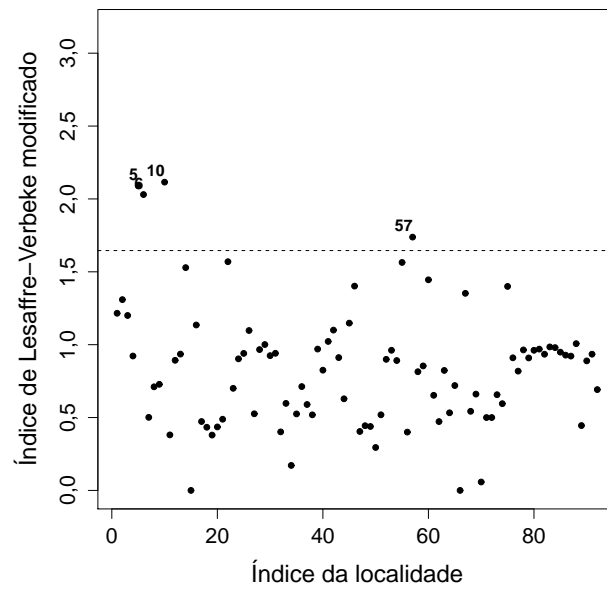


Figura 5.42. Gráfico do índice de Lesaffre-Verbeke modificado *versus* os índices das localidades, modelo (5.28).

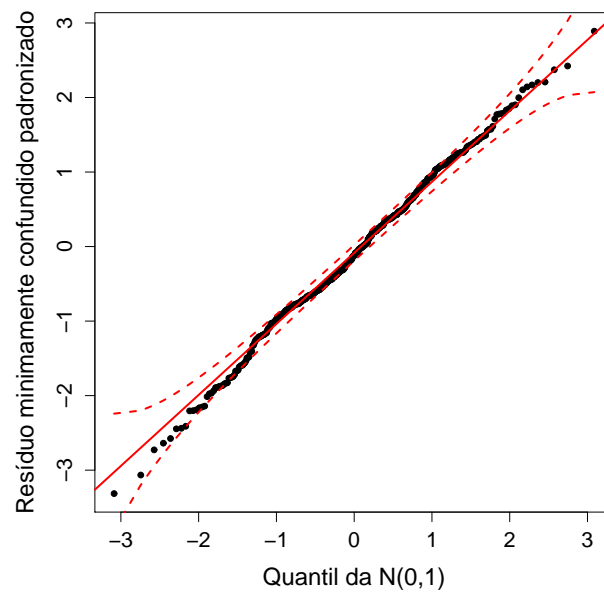


Figura 5.43. Gráfico de probabilidades normal, com envelope gerado, para os resíduos minimamente confundidos padronizados, modelo (5.28).

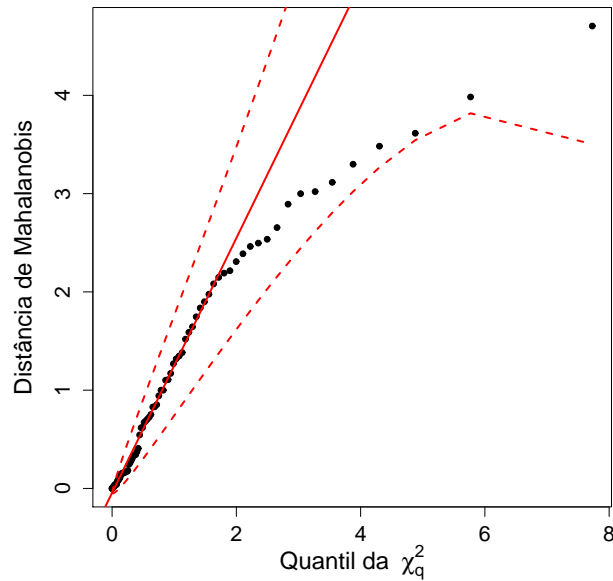


Figura 5.44. Gráficos das distâncias de Mahalanobis, modelo (5.28).

5.4.4.4 Etapa 4: Redução do modelo

Voltamos a investigar a estrutura de média a fim de verificar se as variáveis regressoras e potências selecionadas no modelo (5.24) sofrem alterações uma vez que, agora, a estrutura de covariância é mais complexa. Essa investigação foi feita por meio do algoritmo MFP(0,05; 0,05) com 4 graus de liberdade para cada variável regressora contínua.

Dessa vez, para determinar a ordem de visitação das variáveis regressoras foi utilizada a estatística F , dada a necessidade de utilização da significância conjunta para a variável regressora RM, que entrou no modelo (5.24) com uma função FP2(0; 0,5). A Tabela 5.17 apresenta o valor-p para a determinação da ordem de visitação das variáveis regressoras em todos os ciclos do procedimento MFP.

Podemos observar que a variável regressora RM foi a que apresentou maior efeito e, portanto, será a primeira a ser visitada e a variável regressora TAX será a última. Foram necessários somente 2 ciclos para convergência. O primeiro ciclo do procedimento MFP para os dados das habitações de Boston é apresentado na Tabela

5.18.

Tabela 5.17. Graus de liberdade, estatística F e valor-p ordenados para os parâmetros de regressão do modelo (5.28)

Efeito	Grau de liberdade do numerador	Grau de liberdade do denominador	Estatística F	P
(log(RM), RM ^{0,5})	2	407	227,813	4,10e-67
LSTAT ^{0,5}	1	407	225,904	6,29e-41
OMEGABK	1	407	44,471	8,44e-11
CRIM ^{0,5}	1	407	33,831	1,22e-08
RAD	1	88	37,791	2,25e-08
PTRATIO	1	88	31,128	2,63e-07
DIS	1	407	16,090	7,19e-05
NOX	1	407	15,285	0,0001
TAX	1	88	14,511	0,0003

No ciclo 1 (Tabela 5.18) temos a seleção das mesmas variáveis regressoras da Tabela (5.14), com mudança de potência para a variável regressora RM, de (0; 0,5) para (-0,5; -0,5), e de função para a variável regressora TAX, de linear para uma função FP1 com potência -2. Este ciclo finaliza com $-2LL = -841,493$.

No ciclo 2 (Tabela 5.19) não ocorre nenhuma mudança adicional nas variáveis regressoras ou funções selecionadas, portanto o procedimento MFP finaliza com $-2LL = -841,493$ e com as seguintes variáveis regressoras e suas respectivas transformações: RM^{-0,5}, RM^{-0,5} ln(RM), LSTAT^{0,5}, OMEGABK, CRIM^{0,5}, RAD, PTRATIO, DIS, NOX e TAX⁻². Assim, o modelo PFMM2 fica expresso por

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_{0i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 92, \\ b_{0i} &\sim \mathcal{N}(0, \sigma_0^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2 \delta_{si}^2 \mathbf{R}^*), \quad s = 1, \dots, 24, \end{aligned} \tag{5.29}$$

com

$$\mathbf{X}_i = \begin{bmatrix} 1 & \text{RM}_{i1}^{-0,5} & (\text{RM}^{-0,5} \ln(\text{RM}))_{i1} & \text{LSTAT}_{i1}^{0,5} & \text{OMEGABK}_{i1} & \text{CRIM}_{i1}^{0,5} & \dots & \text{TAX}_{i1}^2 \\ 1 & \text{RM}_{i2}^{-0,5} & (\text{RM}^{-0,5} \ln(\text{RM}))_{i2} & \text{LSTAT}_{i2}^{0,5} & \text{OMEGABK}_{i2} & \text{CRIM}_{i2}^{0,5} & \dots & \text{TAX}_{i2}^2 \\ 1 & \text{RM}_{i3}^{-0,5} & (\text{RM}^{-0,5} \ln(\text{RM}))_{i3} & \text{LSTAT}_{i3}^{0,5} & \text{OMEGABK}_{i3} & \text{CRIM}_{i3}^{0,5} & \dots & \text{TAX}_{i3}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \text{RM}_{i,n_i}^{-0,5} & (\text{RM}^{-0,5} \ln(\text{RM}))_{i,n_i} & \text{LSTAT}_{i,n_i}^{0,5} & \text{OMEGABK}_{i,n_i} & \text{CRIM}_{i,n_i}^{0,5} & \dots & \text{TAX}_{i,n_i}^2 \end{bmatrix},$$

para $s > 1$, $\delta_{si} > 0$, com $i = 4, 15, 28, \dots, 87, 90, 91$, e $\delta_{1i} = 1$, caso contrário; e \mathbf{R}^* definida em (5.27). As estimativas dos parâmetros do modelo (5.29) também são apresentadas na Tabela 5.20.

Para o modelo (5.29) os valores do *AIC*, *BIC* e log-verossimilhança foram $-727,816$, $-572,248$ e $400,908$, respectivamente, em contraste com $-679,328$, $-523,759$ e $376,664$ para o modelo 5.28, confirmando o melhor ajuste para o modelo (5.29) e a necessidade de se visitar a modelagem da função de média após selecionar a função de covariância.

Tabela 5.18. Ciclo 1 do algoritmo MFP para redução do modelo

Variável regressora	Modelo	(vs.)	-2LL	Estatística de teste	P	Potência	(vs.)
RM	Nulo	PFMM2	-549,5364	283,2817	<0,0001*	.	-0,5 ; -0,5
	MLM		-773,1709	59,6472	<0,0001 ⁺	1	
	PFMM1		-811,4307	21,3874	<0,0001 ⁺	3	
	Final		-832,8181			-0,5; -0,5	
LSTAT	Nulo	PFMM2	-657,8964	176,2054	<0,0001*	.	-0,5; 1
	MLM		-821,7552	12,3466	0,0063 ⁺	1	
	PFMM1		-832,8181	1,2837	0,5263	0,5	
	Final		-832,8181			0,5	
OMEGABK	Nulo	PFMM2	-795,4492	39,0761	<0,0001*	.	3; 3
	MLM		-832,8181	1,7071	0,6354	1	
	Final		-832,8181			1	
CRIM	Nulo	PFMM2	-805,3364	31,9175	<0,0001*	.	1; 1
	MLM		-827,3410	9,9129	0,0193 ⁺	1	
	PFMM1		-832,8181	4,4358	0,1088	0,5	
	Final		-832,8181			0,5	
RAD	Nulo	PFMM2	-799,2562	37,2736	<0,0001*	.	-1; 3
	MLM		-832,8181	3,7117	0,2943	1	
	Final		-832,8181			1	
PTRATIO	Nulo	PFMM2	-806,7840	29,9832	<0,0001*	.	0,5; 1
	MLM		-832,8181	3,9490	0,2670	1	
	Final		-832,8181			1	
DIS	Nulo	PFMM2	-818,5295	17,3930	0,0016*	.	-1; -0,5
	MLM		-832,8181	3,1044	0,3758	1	
	Final		-832,8181			1	
NOX	Nulo	PFMM2	-819,8550	15,6948	0,0035*	.	3; 3
	MLM		-832,8181	2,7316	0,4349	1	
	Final		-832,8181			1	
TAX	Nulo	PFMM2	-819,5342	23,1176	0,0001*	.	-2 ; -2
	MLM		-832,8181	9,8337	0,0200 ⁺	1	
	PFMM1		-841,4934	1,1584	0,5604	-2	
	Final		-841,4934			-2	

* Teste para a inclusão de uma variável (binária, categórica ou contínua) é significativo com nível α_1

⁺ Teste PF2 vs PF1 ou PF2 vs linear para a inclusão de uma variável contínua é significativo com nível α_2 .

Tabela 5.19. Ciclo 2 do algoritmo MFP para redução do modelo

Variável regressora	Modelo	(vs.)	-2LL	Estatística de teste	P	Potência	(vs.)
RM	Nulo	PFMM2	-559,4402	282,0532	<0,0001*	.	-0,5; -0,5
	MLM		-781,5211	59,9723	<0,0001 ⁺	1	
	PFMM1		-820,0816	21,4119	<0,0001 ⁺	3	
	Final		-841,4934			-0,5; -0,5	
LSTAT	Nulo	PFMM2	-662,0225	180,5615	<0,0001*	.	-0,5; 1
	MLM		-831,6620	10,9220	0,0122 ⁺	1	
	PFMM1		-841,4766	1,1074	0,5748	0,5	
	Final		-841,4766			0,5	
OMEGABK	Nulo	PFMM2	-804,1180	38,9235	<0,0001*	.	3; 3
	MLM		-841,4934	1,5481	0,6712	1	
	Final		-841,4934			1	
CRIM	Nulo	PFMM2	-813,1392	34,0755	<0,0001*	.	1; 1
	MLM		-836,0542	11,1606	0,0109 ⁺	1	
	PFMM1		-841,4934	5,7213	0,0572	0,5	
	Final		-841,4934			0,5	
RAD	Nulo	PFMM2	-802,8339	40,2844	<0,0001*	.	-0,5; 3
	MLM		-841,4934	1,6249	0,6538	1	
	Final		-841,4934			1	
PTRATIO	Nulo	PFMM2	-812,9955	32,6331	<0,0001*	.	1; 1
	MLM		-841,4934	4,1351	0,2472	1	
	Final		-841,4934			1	
DIS	Nulo	PFMM2	-826,8466	17,8559	0,0013*	.	-0,5; -0,5
	MLM		-841,4934	3,2091	0,3605	1	
	Final		-841,4934			1	
NOX	Nulo	PFMM2	-827,4443	16,5274	0,0024*	.	3; 3
	MLM		-841,4934	2,4783	0,4792	1	
	Final		-841,4934			1	
TAX	Nulo	PFMM2	-819,5342	23,1176	0,0001*	.	-2; -2
	MLM		-832,8181	9,8337	0,0200 ⁺	1	
	PFMM1		-841,4934	1,1584	0,5604	-2	
	Final		-841,4934			-2	

* Teste para a inclusão de uma variável (binária, categórica ou contínua) é significativo com nível α_1

⁺ Teste PFMM2 vs PFMM1 ou PFMM2 vs MLM para a inclusão de uma variável contínua é significativo com nível α_2 .

Tabela 5.20. Estimativas de máxima verossimilhança restrita e erros padrão estimados associados aos parâmetros dos modelos (5.23), (5.24), (5.28) e (5.29)

Parâmetros dos efeitos fixos	PFMM							
	Modelo misto		Etapa 1		Etapa 3		Etapa 4	
	Estimativa	E.P.	Estimativa	E.P.	Estimativa	E.P.	Estimativa	E.P.
Intercepto	3,176	0,267	2,144	0,358	0,331	0,247	29,733	1,995
CRIM	-0,006	0,001	—	—	—	—	—	—
CRIM ^{0,5}	—	—	-0,075	0,010	-0,075	0,013	-0,080	0,013
ZN	0,001	0,001	—	—	—	—	—	—
INDUS	0,004	0,005	—	—	—	—	—	—
CHAS	-0,019	0,028	—	—	—	—	—	—
NOX	-0,870	0,164	-0,767	0,157	-0,498	0,127	-0,499	0,125
RM	0,143	0,014	—	—	—	—	—	—
ln(RM)	—	—	-6,114	1,105	-9,748	1,067	—	—
RM ^{0,5}	—	—	5,433	0,890	8,635	0,835	—	—
RM ^{-0,5}	—	—	—	—	—	—	-10,508	0,505
RM ^{-0,5} ln(RM)	—	—	—	—	—	—	-29,622	2,520
AGE	-0,001	0,000	—	—	—	—	—	—
DIS	-0,039	0,011	-0,031	0,008	-0,018	0,005	-0,018	0,004
RAD	0,020	0,005	0,022	0,004	0,019	0,003	0,014	0,002
TAX	-0,001	0,000	-0,001	0,000	-0,001	0,000	—	—
TAX ⁻²	—	—	—	—	—	—	0,797	0,163
PTRATIO	-0,029	0,011	-0,027	0,009	-0,025	0,005	-0,025	0,004
BK	0,481	0,076	—	—	—	—	—	—
OMEGABK	—	—	0,313	0,057	0,345	0,052	0,346	0,052
LSTAT	-0,020	0,002	—	—	—	—	—	—
LSTAT ^{0,5}	—	—	-0,183	0,012	-0,132	0,009	-0,131	0,009
Componentes de variância								
Intercepto	0,024	—	0,017	—	5,30e-08	—	4,67e-07	—
Resíduo								
Único grupo	0,016	—	0,016	—	8,05e-03	—	7,52e-03	—
Grupo 1	—	—	—	—	1,000	—	1,000	—
Grupo 2	—	—	—	—	3,542	—	3,686	—
Grupo 3	—	—	—	—	4,475	—	3,177	—
Grupo 4	—	—	—	—	2,226	—	2,283	—
Grupo 5	—	—	—	—	3,440	—	3,589	—
Grupo 6	—	—	—	—	3,474	—	3,046	—
Grupo 7	—	—	—	—	2,681	—	2,618	—
Grupo 8	—	—	—	—	3,145	—	3,458	—
Grupo 9	—	—	—	—	2,096	—	2,012	—
Grupo 10	—	—	—	—	2,465	—	2,388	—
Grupo 11	—	—	—	—	10,374	—	12,822	—
Grupo 12	—	—	—	—	55,858	—	57,793	—
Grupo 13	—	—	—	—	47,884	—	52,401	—
Grupo 14	—	—	—	—	5,807	—	6,219	—
Grupo 15	—	—	—	—	10,593	—	10,886	—
Grupo 16	—	—	—	—	16,676	—	17,138	—
Grupo 17	—	—	—	—	14,926	—	17,299	—
Grupo 18	—	—	—	—	10,094	—	10,748	—
Grupo 19	—	—	—	—	2,206	—	2,221	—
Grupo 20	—	—	—	—	3,392	—	3,560	—
Grupo 21	—	—	—	—	2,407	—	2,682	—
Grupo 22	—	—	—	—	5,915	—	5,963	—
Grupo 23	—	—	—	—	20,123	—	22,413	—
Grupo 24	—	—	—	—	4,752	—	5,083	—
ρ	—	—	—	—	0,455	—	0,430	—
<i>AIC</i>	-326,473	—	-423,686	—	-679,328	—	-709,396	—
<i>BIC</i>	-259,298	—	-369,026	—	-523,759	—	-553,827	—
Log-verosim	179,237	—	224,843	—	376,664	—	391,698	—

Os gráficos dos resíduos do modelo (5.29) estão apresentados nas Figuras 5.45–5.50.

Analisando a Figura 5.45, vemos padrão condizente com hipótese de linearidade para os efeitos fixos, pois não se observa nenhum tipo de tendência do resíduo marginal. Uma análise da Figura 5.46 mostra que a localidade 39 destaca levemente o setor censitário 10 como discrepante, porém com valor do resíduo sobre a faixa da banda de confiança.

Pela Figura 5.47, não se percebe nenhum afastamento da suposição de normalidade do erro condicional por parte do resíduo com confundimento mínimo.

O gráfico para a estrutura de covariância, Figura 5.48, indica que a estrutura de covariância adotada no modelo (5.29) está adequada.

Na Figura 5.49, o gráfico da distância de Mahalanobis não identifica nenhuma localidade como considerada discrepante. Na Figura 5.50, o gráfico dos quantis baseado na distribuição qui-quadrado para essas distâncias indica ser plausível os efeitos aleatórios seguirem uma distribuição gaussiana.

Gráficos com valores da distância de Cook condicional por setor censitário e seus respectivos valores decompostos são apresentados nas Figuras 6.15–6.18 do Anexo. Pode-se verificar que, para todas as observações, os valores da distância se concentram em torno do zero, não sendo, portanto, nenhuma observação classificada como possível ponto influente nas estimativas dos efeitos fixos e/ou aleatórios.

Essa aplicação ilustra um caso de múltiplas variáveis regressoras contínuas na parte fixa de um modelo misto, no qual há indícios de não linearidade nos efeitos de algumas delas. Suspeita sobre esse problema foi encontrado no estudo inicial de Harrison & Rubinfeld (1978), que incluíram em seu modelo as variáveis RM, DIS, RAD, LSTAT, NOX transformadas. A estratégia de modelagem proposta permitiu a investigação da falta de linearidade no conjunto de variáveis regressoras e possibilitou um ajuste melhor aos dados.

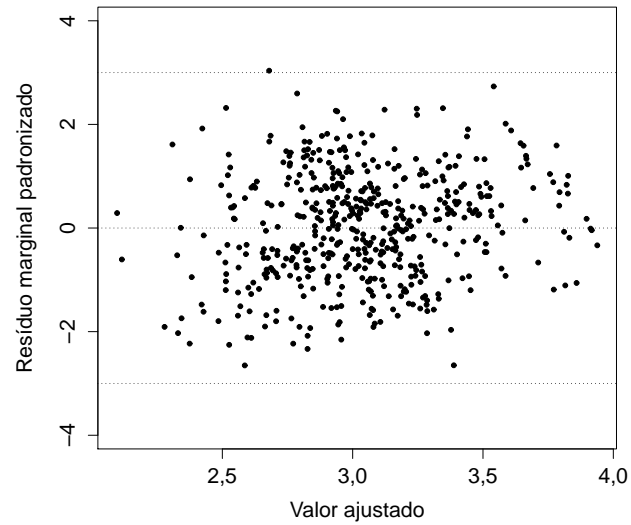


Figura 5.45. Gráfico do resíduo marginal padronizado *versus* valor ajustado para o modelo em (5.29).

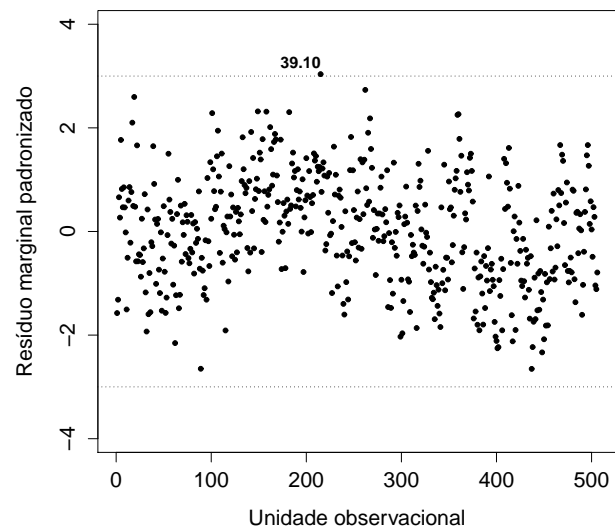


Figura 5.46. Gráfico do resíduo marginal padronizado *versus* unidade observacional para o modelo em 5.29.

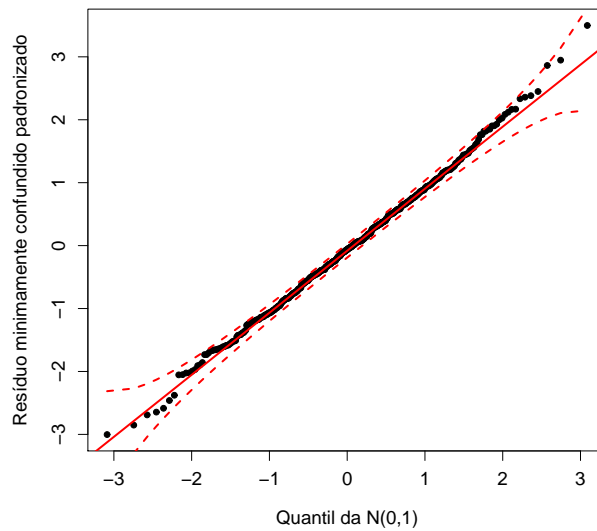


Figura 5.47. Gráfico do envelope simulado com 95% para o resíduo com confundimento mínimo para o modelo em (5.29).

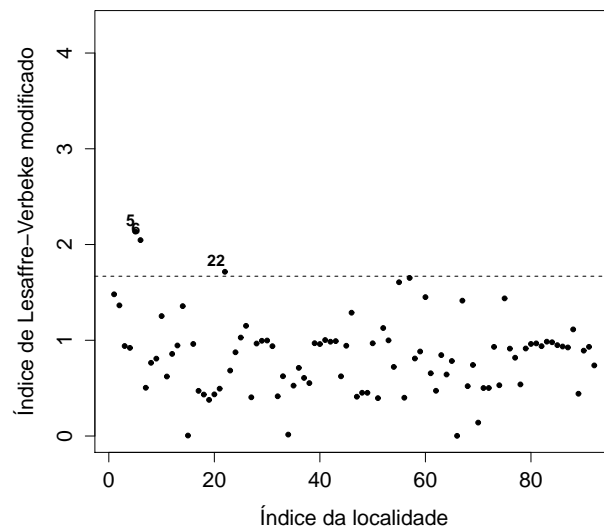


Figura 5.48. Gráfico do índice de Lesaffre-Verbeke modificado *versus* índice da localidade para o modelo em (5.29).

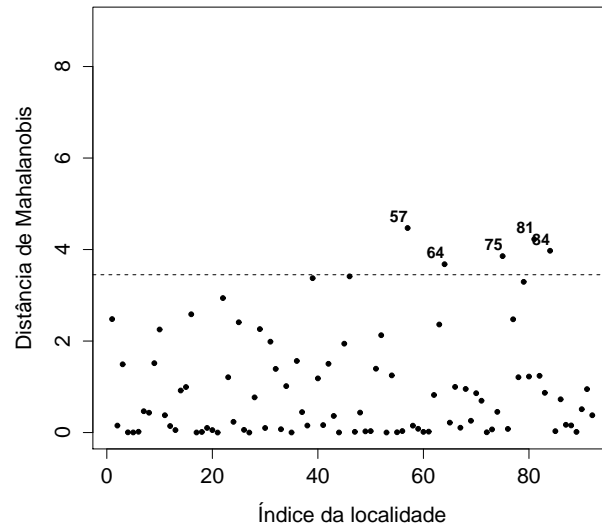


Figura 5.49. Gráficos baseado na distância de Mahalanobis *versus* índice da localidade para o modelo em (5.29).

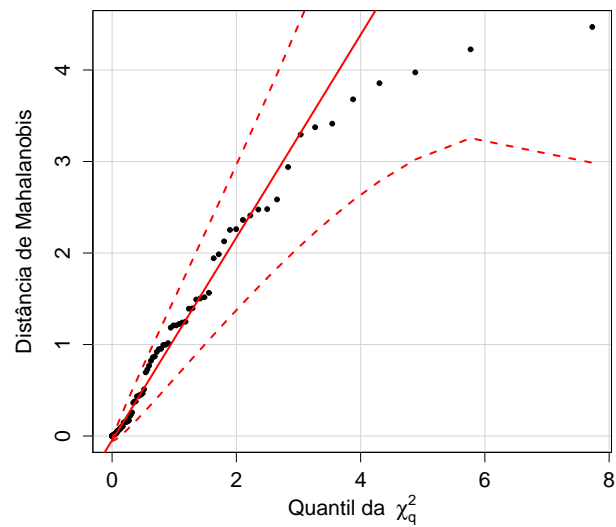


Figura 5.50. Gráficos QQ baseado na distância de Mahalanobis correspondentes ao modelo em (5.29).

6 DISCUSSÕES E CONCLUSÕES

Nesta tese propomos uma estratégia para seleção de polinômios fracionários dentro da classe de modelos lineares mistos na presença de múltiplas variáveis regressoras, inclusive com termos de interação, e modelagem da estrutura de covariância dos dados. A estratégia proposta estende os algoritmos MFP e MFPI na presença de dados agrupados. Aqui vale trazer a reflexão de Royston & Sauerbrei (2008) “Do ponto de vista teórico, qualquer procedimento de construção de modelos pode facilmente ser criticado. No entanto, pesquisadores aplicados esperam e precisam de respostas sobre a importância e forma funcional de variáveis regressoras contínuas”. A construção de modelos, até mesmo sob a suposição de efeitos lineares em regressão múltipla, não é trivial e nem existe um único método de consenso, visto a diversidade de procedimentos e critérios de seleção que abundam na literatura. No caso de modelos mistos, temos a complexidade da modelagem das duas partes envolvidas no modelo. A estratégia explorada é baseada em princípios historicamente recomendados: especificação mais completa possível da parte fixa e hierarquia entre efeitos. Nossa expectativa é de que as ideias aqui colocadas auxiliem a modelagem de efeitos mistos, ampliando as possibilidades de tratamentos de relações não lineares, e sejam de grande utilidade prática. Para isso, desenvolvemos funções para o ajuste do modelo PFMM, fazendo uso do pacote `nlme`.

A estratégia proposta foi aplicada em três conjuntos de dados, ilustrando a utilidade dos FPs. Nas duas primeiras aplicações os FPs se mostraram flexíveis para modelar as relações não lineares no contexto dos modelos lineares mistos. Em um dos casos foi possível obter um ajuste, com apenas dois componentes para a variável regressora, tão bom quanto aquele fornecido por um polinômio convencional

de quarta ordem. No outro exemplo, o modelo PFMM foi muito superior ao modelo polinomial de terceira ordem, resultando numa curva com ponto de máximo, mas assimétrica, um padrão de relação bastante comum em dados longitudinais. Na terceira aplicação foi explorado um caso com múltiplas variáveis regressoras contínuas na qual foi possível encontrar as transformações FP e selecionar variáveis de uma maneira objetiva.

Foi vantajoso o uso do conjunto S no espaço discreto proposto por Royston e coautores já que esse conjunto restrito viabiliza a seleção na presença de muitas variáveis regressoras. No entanto, uma possibilidade seria aumentar a classe de curvas disponíveis, acrescentando outros valores ao conjunto para a busca das transformações FP, o que pode ser recomendado para algumas aplicações com número não muito excessivo de variáveis regressoras. Existe também a possibilidade de se ampliar o espaço paramétrico das potências para o espaço contínuo, o que pode ser promissor para poucas variáveis regressoras. Para várias variáveis regressoras a expectativa é a de que o método não seja de muito sucesso devido aos problemas de convergências em otimização não linear, além da perda da natureza sistemática de visitação das variáveis regressoras, oferecida pelo algoritmo MFP.

As funções desenvolvidas e utilizadas podem ser solicitadas diretamente via e-mail: edijane@ufam.edu.br.

6.1 Pesquisas futuras

Defendemos a ideia de que os FPs podem ser incorporados em qualquer modelo que, em alguma parte, envolva uma expressão que remeta à forma de um preditor linear. Assim, as possibilidades de extensão da pesquisa são vastas e, dentre elas, destacamos:

- Estender o modelo PFMM para interações entre variáveis regressoras contínuas através de adaptações no algoritmo MFPIGen (*Multivariable Fractional Polynomials Interaction General*) (Royston & Sauerbrei, 2008);

- Propor estratégias de seleção para polinômios fracionários dentro da classe de modelos lineares generalizados mistos;
- Abordar o problema de não linearidade nos modelos com erros de medida na covariável.

Anexos

Anexo 1 - Gráficos de diagnóstico para o modelo em (5.8)

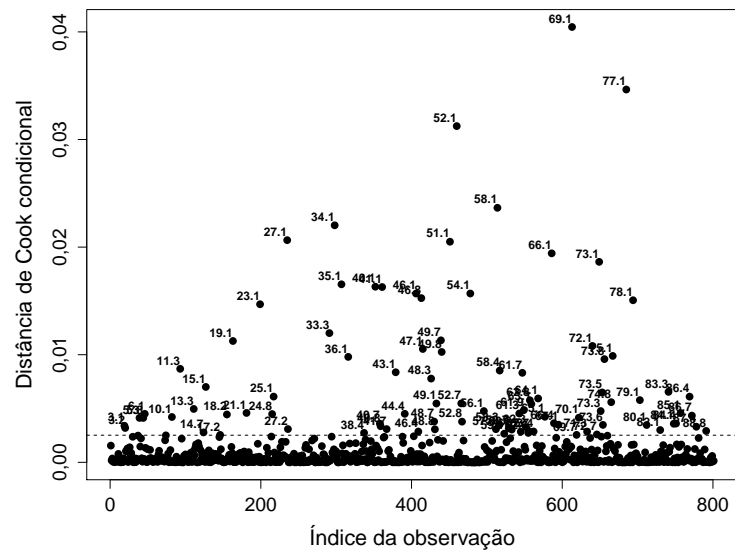


Figura 6.1. Gráfico da distância de Cook condicional para o modelo (5.8).

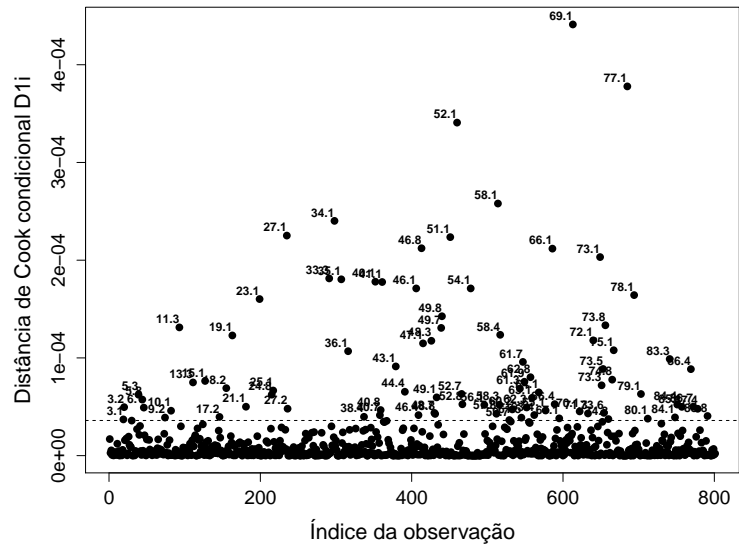


Figura 6.2. Gráfico da distância de Cook condicional $D1i$ para o modelo (5.8).

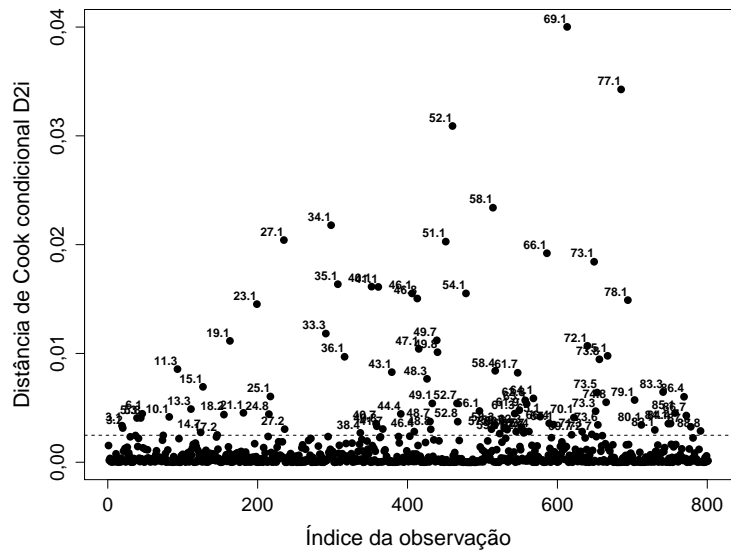


Figura 6.3. Gráfico da distância de Cook condicional $D2i$ para o modelo (5.8).

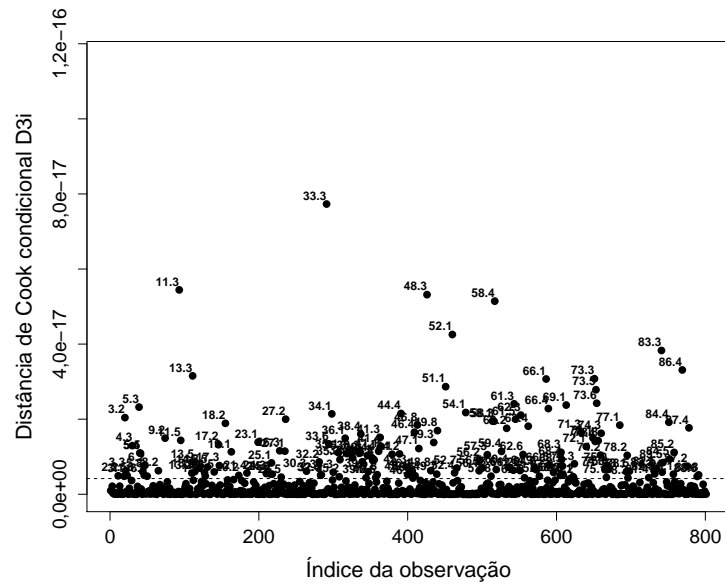


Figura 6.4. Gráfico da distância de Cook condicional D_{3i} para o modelo (5.8).

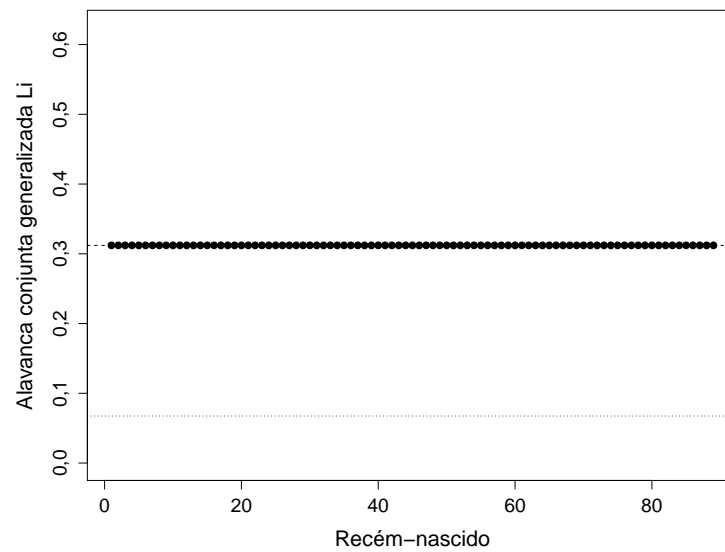


Figura 6.5. Gráfico da alavanca marginal generalizada L_i para o modelo (5.8).

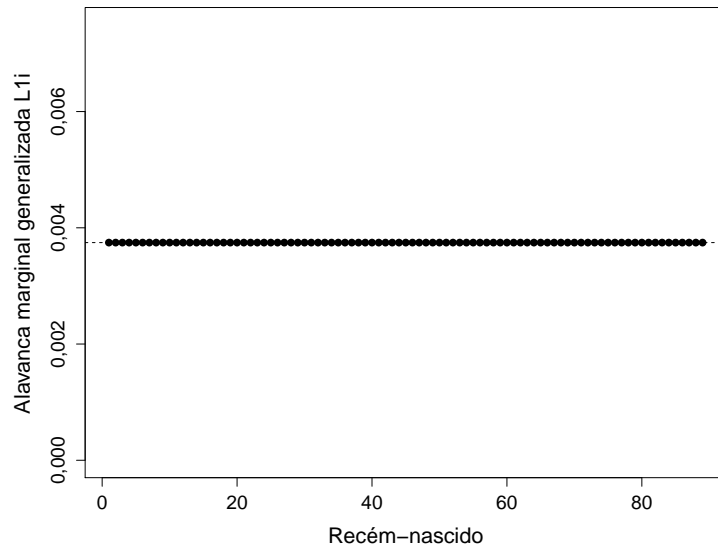


Figura 6.6. Gráfico da alavanca marginal generalizada L1i para o modelo (5.8).

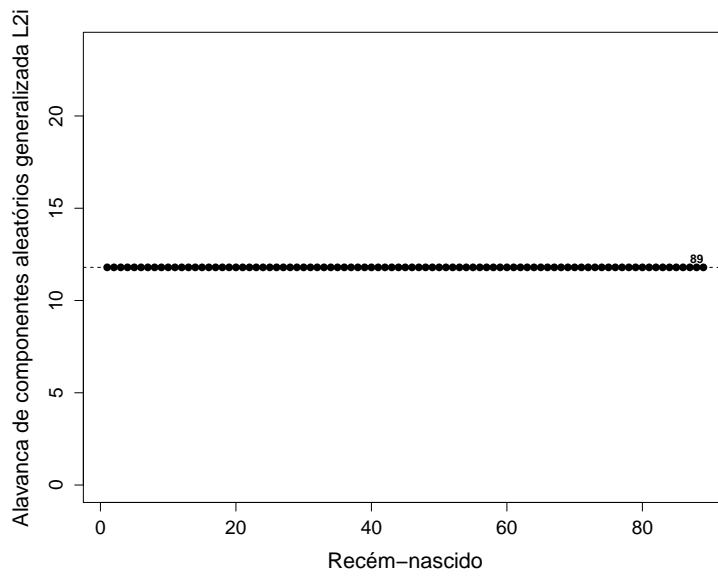


Figura 6.7. Gráfico da alavanca marginal generalizada L2i para o modelo (5.8).

Anexo 2 - Gráficos de diagnóstico para o modelo em (5.20)

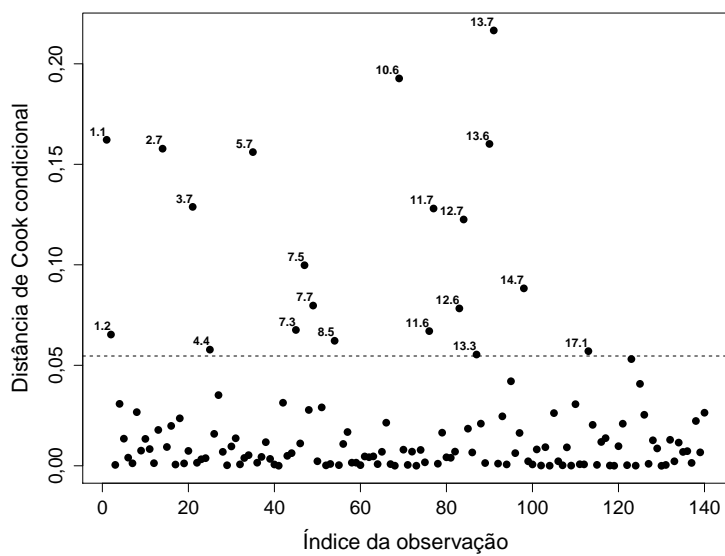


Figura 6.8. Gráfico da distância de Cook condicional para o modelo (5.20).

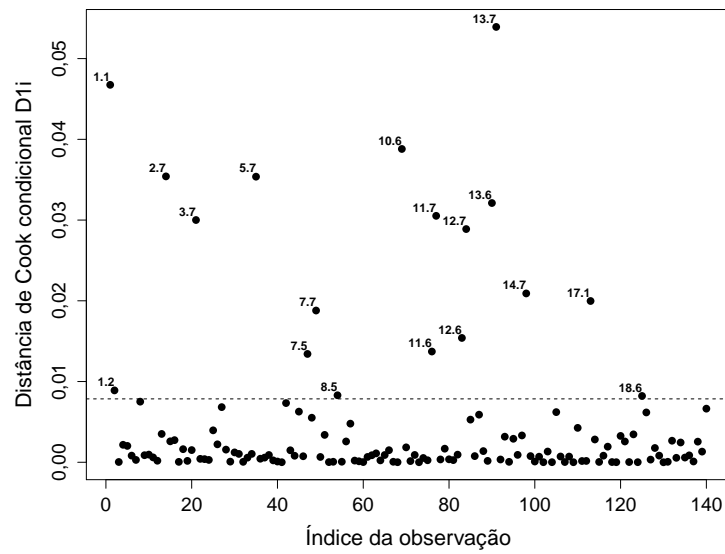


Figura 6.9. Gráfico da distância de Cook condicional $D1i$ para o modelo (5.20).

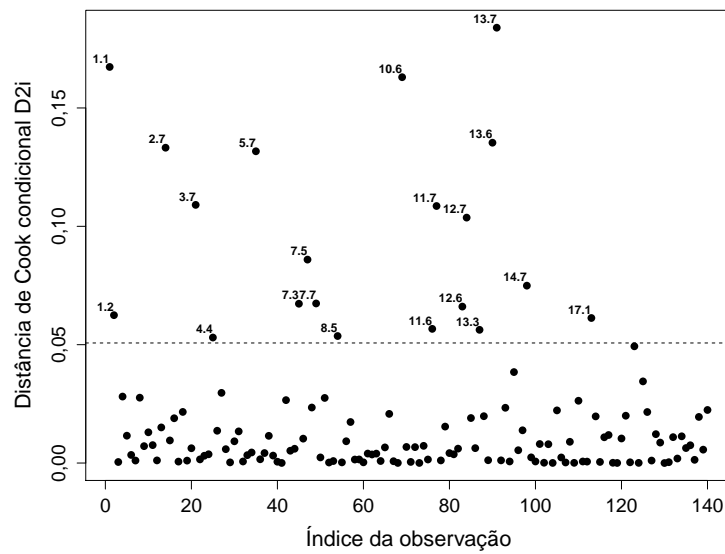


Figura 6.10. Gráfico da distância de Cook condicional $D2i$ para o modelo (5.20).

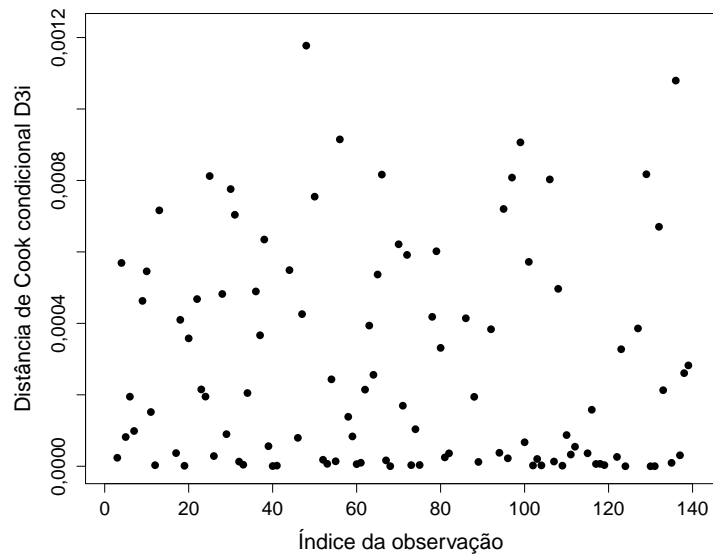


Figura 6.11. Gráfico da distância de Cook condicional D_{3i} para o modelo (5.20).

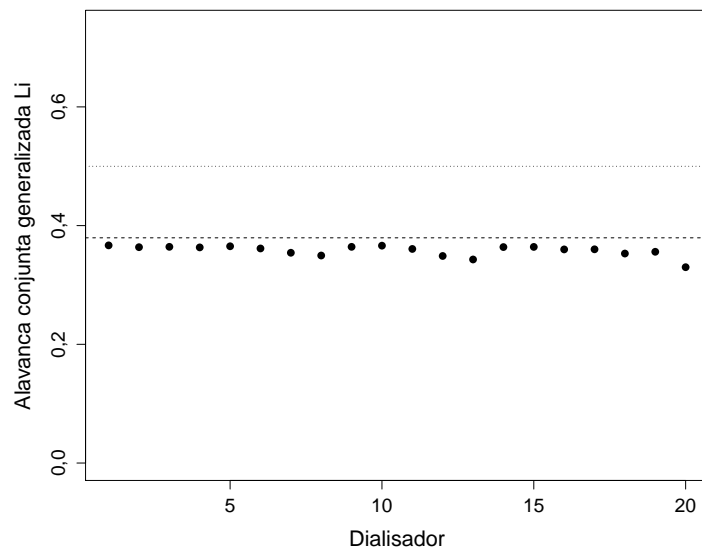


Figura 6.12. Gráfico da alavanca marginal generalizada L_i para o modelo (5.20).

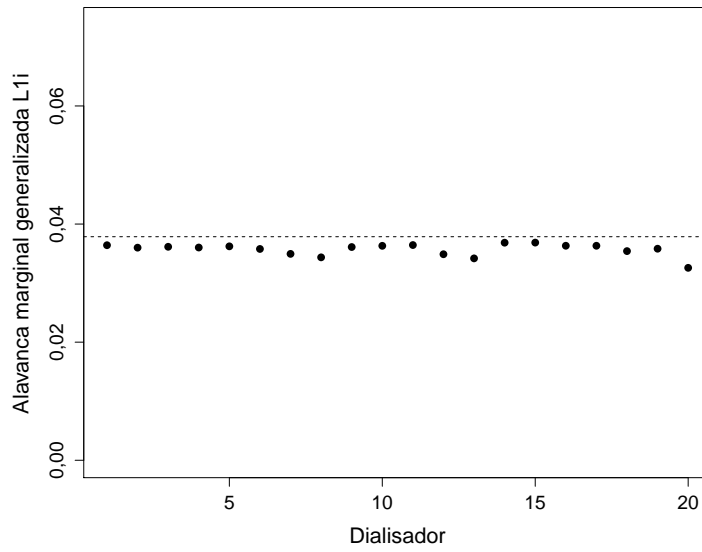


Figura 6.13. Gráfico da alavanca marginal generalizada L1i para o modelo (5.20).

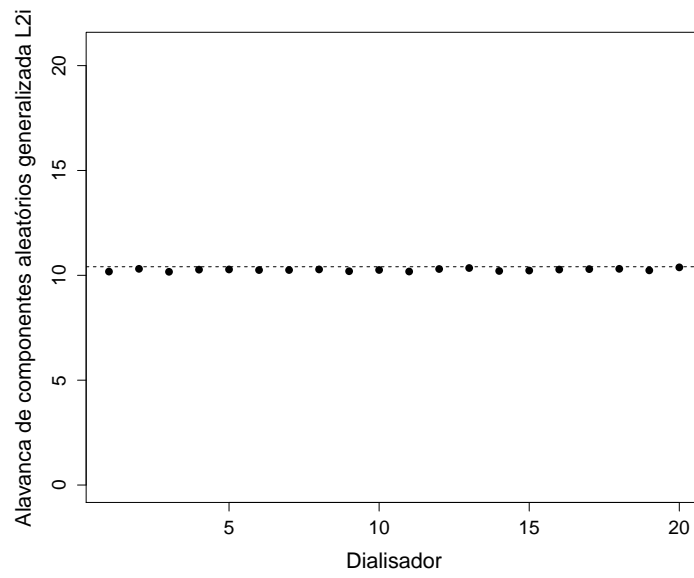


Figura 6.14. Gráfico da alavanca marginal generalizada L2i para o modelo (5.20).

Anexo 3 - Gráficos de diagnóstico para o modelo em (5.29)

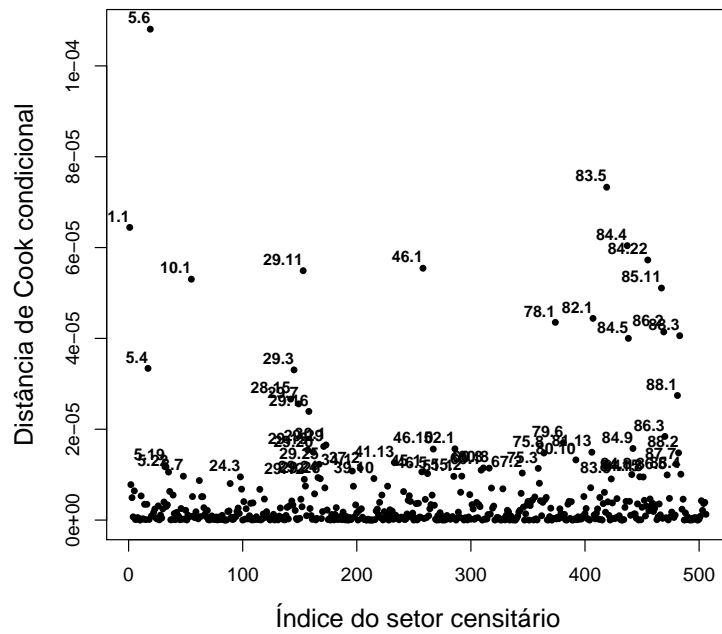


Figura 6.15. Gráfico da distância de Cook condicional para o modelo (5.29).

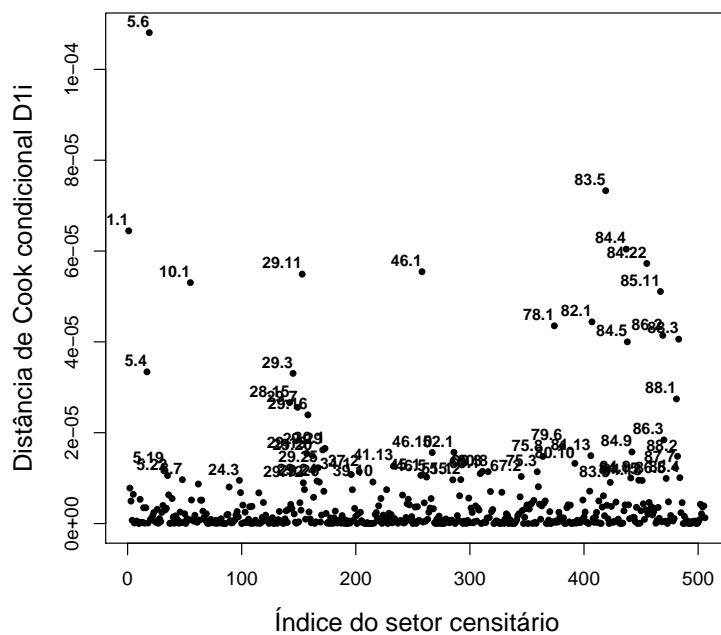


Figura 6.16. Gráfico da distância de Cook condicional D1i para o modelo (5.29).

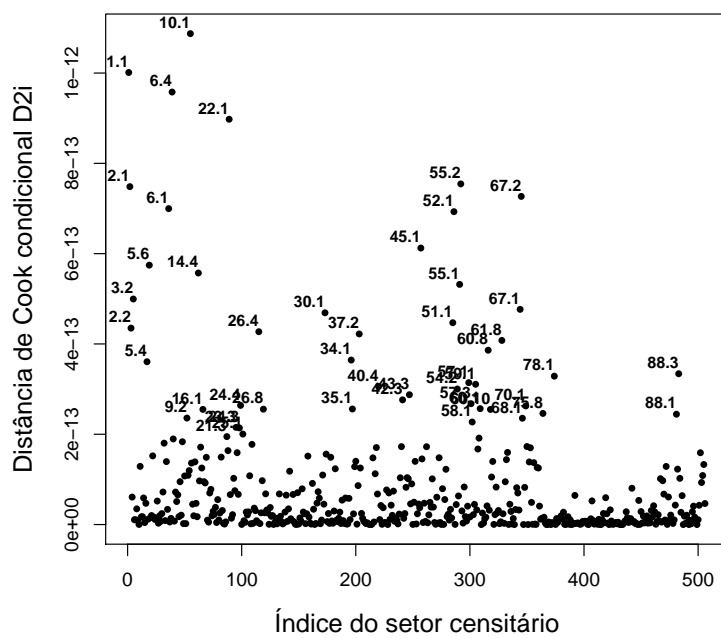


Figura 6.17. Gráfico da distância de Cook condicional D2i para o modelo (5.29).

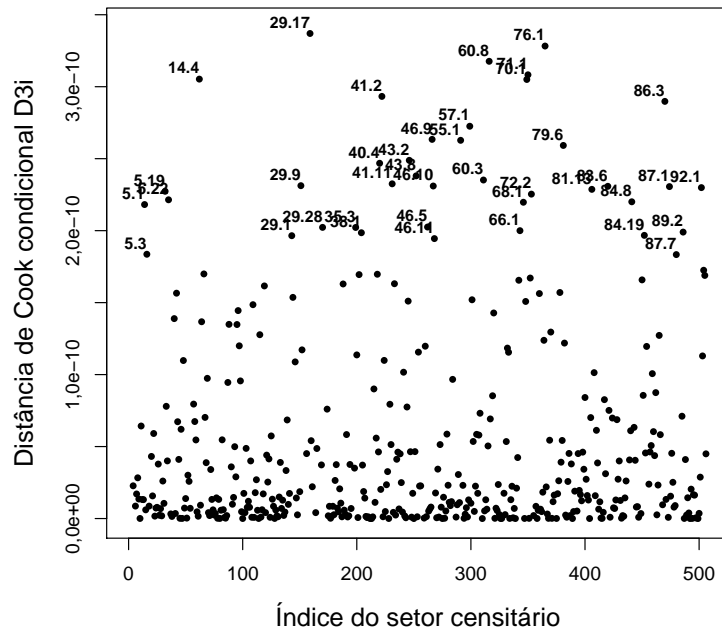


Figura 6.18. Gráfico da distância de Cook condicional D_{3i} para o modelo (5.29).

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716–723, 1974.

ALTMAN, D. G.; LAUSEN, B.; SAUERBREI, W.; SCHUMACHER, M. Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. **JNCI: Journal of the National Cancer Institute**, v. 86, n. 11, p. 829–835, 1994.

AMBLER, G.; ROYSTON, P. Fractional polynomial model selection procedures: investigation of Type I error rate. **Journal of Statistical Computation and Simulation**, v. 69, n. 1, p. 89–108, 2001.

AREGAY, M.; SHKEDY, Z.; MOLENBERGHS, G.; DAVID, M.-P.; TIBALDI, F. Model-based estimates of long-term persistence of induced HPV antibodies: a flexible subject-specific approach. **Journal of Biopharmaceutical Statistics**, v. 23, n. 6, p. 1228–1248, 2013.

ARELLANO-VALLE, R. B.; BOLFARINE, H.; LACHOS, V. H. Skew-normal linear mixed models. **Journal of Data Science**, v. 3, n. 4, p. 415–438, 2005.

AUSTIN, P. C.; BRUNNER, L. J. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. **Statistics in medicine**, v. 23, n. 7, p. 1159–1178, 2004.

AUSTIN, P. C.; PARK-WYLLIE, L. Y.; JUURLINK, D. N. Using fractional polynomials to model the effect of cumulative duration of exposure on outcomes: applications to cohort and nested case-control designs. **Pharmacoepidemiology and Drug Safety**, v. 23, n. 8, p. 819–829, 2014.

BAGNARDI, V.; ZAMBON, A.; QUATTO, P.; CORRAO, G. Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. **American journal of epidemiology**, v. 159, n. 11, p. 1077–1086, 2004.

BANESHI, M. R.; NAKHAE, F.; LAW, M. On the use of fractional polynomial models to assess preventive aspect of variables: An example in prevention of mortality following HIV infection. **International Journal of Preventive Medicine**, v. 4, n. 4, p. 414–419, 2013.

BATES, D.; MÄCHLER, M.; BOLKER, B.; WALKER, S. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1–48, 2015.

BATES, D. M.; WATTS, D. G. **Nonlinear Regression Analysis and Its Applications**. New York: John Wiley & Sons, 1988. 365 p.

BECKMAN, R. J.; NACHTSHEIM, C. J.; COOK, R. D. Diagnostics for mixed-model analysis of variance. **Technometrics**, v. 29, n. 4, p. 413–426, 1987.

BELSLEY, D. A.; KUH, E.; WELSCH, R. E. **Regression Diagnostics: Identifying Influential Data and Sources of Collinearity**. New York: John Wiley & Sons, 1980. 292 p.

BERGER, U.; SCHÄFER, J.; ULM, K. Dynamic Cox modelling based on fractional polynomials: time-variations in gastric cancer prognosis. **Statistics in Medicine**, v. 22, n. 7, p. 1163–1180, 2003.

BOLLAERTS, K.; AERTS, M.; FAES, C.; GRIJSPEERDT, K.; DEWULF, J.; MINTIENS, K. Human salmonellosis: Estimation of dose-illness from outbreak data. **Risk Analysis**, v. 28, n. 2, p. 427–440, 2008.

BOSSARD, N.; DESCOTES, F.; BREMOND, A. G.; BOBIN, Y.; DE SAINT HILAIRE, P.; GOLFIER, F.; AWADA, A.; MATHEVET, P. M.; BERRERD, L.; BARBIER, Y.; ESTÈVES, J. Keeping data continuous when analyzing the prognostic impact of a tumor marker: an example with cathepsin D in breast cancer. **Breast cancer research and treatment**, v. 82, n. 1, p. 47–59, 2003.

BOX, G. E. P.; TIDWELL, P. W. Transformation of the independent variables. **Technometrics**, v. 4, n. 4, p. 531–550, 1962.

BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, n. 3, p. 345–370, 1987.

BREIMAN, L.; FRIEDMAN, J. H. Estimating optimal transformations for multiple regression and correlation. **Journal of the American Statistical Association**, v. 80, n. 391, p. 580–598, 1985.

CHATTERJEE, S.; HADI, A. S. **Sensitivity Analysis in Linear Regression**. John Wiley & Sons, 1988. 315 p.

COOK, R. D. Detection of influential observation in linear regression. **Technometrics**, v. 19, n. 1, p. 15–18, 1977.

COOK, R. D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society, B**, p. 133–169, 1986.

COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. New York: Chapman and Hall, 1982. 230 p.

COX, D. R.; HINKLEY, D. V. **Theoretical Statistics**. London: Chapman and Hall, 1974. 511 p.

CRAINICEANU, C. M.; RUPPERT, D. Likelihood ratio tests in linear mixed models with one variance component. **Journal of the Royal Statistical Society, B**, v. 66, n. 1, p. 165–185, 2004.

DAVIDIAN, M.; GILTINAN, D. **Nonlinear Models for Repeated Measurement Data**. New York: Chapman & Hall, 1995.

DEMIDENKO, E. **Mixed Models: Theory and Applications with R**. 2. ed. Hoboken, New Jersey: John Wiley & Sons, 2013. 754p.

DEMIDENKO, E.; STUKEL, T. A. Influence analysis for linear mixed-effects models. **Statistics in medicine**, v. 24, n. 6, p. 893–909, 2005.

DI CASTELNUOVO, A.; COSTANZO, S.; BAGNARDI, V.; DONATI, M. B.; IACOVIELLO, L.; DE GAETANO, G. Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies. **Archives of Internal Medicine**, v. 166, n. 22, p. 2437–2445, 2006.

DIAS, R. **Regressão Não Paramétrica**, 2001. Disponível em: <<http://www.ime.unicamp.br/~dias/np.html>>. Acesso em: 02 jan. 2018.

DIGGLE, P. J.; HEAGERTY, P. J.; LIANG, K.-Y.; ZEGER, S. L. **Analysis of Longitudinal Data**. 2. ed. United Kingdom: Oxford University Press, 2002. 379p.

DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. John Wiley & Sons, 1998. 706v.

DRIKVANDI, R.; VERBEKE, G.; KHODADADI, A.; NIA, V. P. Testing multiple variance components in linear mixed-effects models. **Biostatistics**, v. 14, n. 1, p. 144–159, 2013.

FAES, C.; AERTS, M.; GEYS, H.; MOLENBERGHS, G. Model averaging using fractional polynomials to estimate a safe level of exposure. **Risk Analysis**, v. 27, n. 1, p. 111–123, 2007.

FAES, C.; GEYS, H.; AERTS, M.; MOLENBERGHS, G. Use of fractional polynomials for dose-response modelling and quantitative risk assessment in developmental toxicity studies. **Statistical Modelling**, v. 3, n. 2, p. 109–125, 2003.

FAES, C.; HENS, N.; AERTS, M.; SHKEDY, Z.; GEYS, H.; MINTIENS, K.; LAEVENS, H.; BOELAERT, F. Estimating herd-specific force of infection by using random-effects models for clustered binary data and monotone fractional polynomials. **Journal of the Royal Statistical Society, C**, v. 55, n. 5, p. 595–613, 2006.

- FEI, Y.; PAN, J. Influence assessments for longitudinal data in linear mixed models. In: **18th International Workshop on Statistical Modelling**. Leuven: Katholieke Universiteit Leuven, 2003. 143–148.
- FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**. 2. ed. Thousand Oaks CA: Sage, 2011.
- FUNG, W.-K.; ZHU, Z.-Y.; WEI, B.-C.; HE, X. Influence diagnostics and outlier tests for semiparametric mixed models. **Journal of the Royal Statistical Society, B**, v. 64, n. 3, p. 565–579, 2002.
- GAŁECKI, A.; BURZYKOWSKI, T. **Linear Mixed-Effects Models Using R: A Step-by-Step Approach**. New York: Springer Science+Business Media, 2013. 542 p.
- ORIGINAL DE GARETH AMBLER; MODIFICADO POR AXEL BENNER. **mfp: Multivariable Fractional Polynomials**, 2015. R package version 1.5.2.
- GEYS, H.; MOLENBERGHS, G.; RYAN, L. M. Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. **Journal of the American Statistical Association**, v. 94, n. 447, p. 734–745, 1999.
- GIAMPAOLI, V.; SINGER, J. M. Likelihood ratio tests for variance components in linear mixed models. **Journal of Statistical Planning and Inference**, v. 139, n. 4, p. 1435–1448, 2009.
- GILLEY, O. W.; PACE, R. K. On the Harrison and Rubinfeld Data. **Journal of Environmental Economics and Management**, v. 31, n. 3, p. 403–405, 1996.
- GILMOUR, S. G.; TRINCA, L. A. Fractional polynomial response surface models. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 10, n. 1, p. 50–60, 2005.

HARRISON, D.; RUBINFELD, D. L. Hedonic housing prices and the demand for clean air. **Journal of Environmental Economics and Management**, v. 5, p. 81–102, 1978.

HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association**, v. 72, n. 358, p. 320–338, 1977.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, p. 423–447, 1975.

HILDEN-MINTON, J. A. Multilevel diagnostics for mixed and hierarchical linear models. Los Angeles, 1995. Tese (Doutorado) - University of California.

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

HOPE, R. M. **Rmisc: Ryan Miscellaneous**, 2013. R package version 1.5.

IBACACHE-PULGAR, G.; PAULA, G. A.; GALEA, M. Influence diagnostics for elliptical semiparametric mixed models. **Statistical Modelling**, v. 12, n. 2, p. 165–193, 2012.

IRWIN, J. R.; MCCLELLAND, G. H. Negative consequences of dichotomizing continuous predictor variables. **Journal of Marketing Research**, v. 40, n. 3, p. 366–371, 2003.

JANSEN, J. P. Network meta-analysis of survival data with fractional polynomials. **BMC medical research methodology**, v. 11, n. 61, p. 1–14, 2011.

JANSEN, J. P.; COPE, S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. **BMC medical research methodology**, v. 12, n. 152, p. 1–16, 2012.

JIANG, J. Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. **Statistica Sinica**, v. 8, n. 3, p. 861–885, 1998.

JIANG, J. On unbiasedness of the empirical BLUE and BLUP. **Statistics & Probability Letters**, v. 41, n. 1, p. 19–24, 1999.

JOHNSON, W.; BALAKRISHNA, N.; GRIFFITHS, P. L. Modeling physical growth using mixed effects models. **American journal of physical anthropology**, v. 150, n. 1, p. 58–67, 2013.

KASENDA, B.; SAUERBREI, W.; ROYSTON, P.; MERCAT, A.; SLUTSKY, A. S.; COOK, D.; GUYATT, G. H.; BROCHARD, L.; RICHARD, J.-C. M.; STEWART, T. E.; MEADE, M.; BRIEL, M. Multivariable fractional polynomial interaction to investigate continuous effect modifiers in a meta-analysis on higher versus lower PEEP for patients with ARDS. **BMJ open**, v. 6, n. 9, 2016.

LACHOS, V. H.; GHOSH, P.; ARELLANO-VALLE, R. B. Likelihood based inference for skew-normal independent linear mixed models. **Statistica Sinica**, v. 20, n. 1, p. 303–322, 2010.

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

LAMBERT, P. C.; SMITH, L. K.; JONES, D. R.; BOTHA, J. L. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. **Statistics in medicine**, v. 24, n. 24, p. 3871–3885, 2005.

LANGE, N.; RYAN, L. Assessing normality in random effects models. **The Annals of Statistics**, v. 17, n. 2, p. 624–642, 1989.

LEE, O. E.; BRAUN, T. M. Permutation tests for random effects in linear mixed models. **Biometrics**, v. 68, n. 2, p. 486–493, 2012.

LEISCH, F.; DIMITRIADOU, E. **mlbench: Machine Learning Benchmark Problems**, 2010. R package version 2.1-1.

LESAFFRE, E.; VERBEKE, G. Local influence in linear mixed models. **Biometrics**, v. 54, n. 2, p. 570–582, 1998.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS System for Mixed Models**. Cary, North Carolina: SAS Institute Inc., 1996.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D.; SCHABENBERGER, O. **SAS for Mixed Models**. 2. ed. Cary, North Carolina: SAS Institute Inc., 2006. 814p.

LONG, J.; RYOO, J. Using fractional polynomials to model non-linear trends in longitudinal data. **British Journal of Mathematical and Statistical Psychology**, v. 63, n. 1, p. 177–203, 2010.

MACCALLUM, R. C.; ZHANG, S.; PREACHER, K. J.; RUCKER, D. D. On the practice of dichotomization of quantitative variables. **Psychological methods**, v. 7, n. 1, p. 19–40, 2002.

MARTINS, A. M.; SILVESTRE, A. M.; PETIM-BATISTA, M. F.; COLAÇO, J. A. Somatic cell score genetic parameter estimates of dairy cattle in Portugal using fractional polynomials 1. **Journal of Animal Science**, v. 89, n. 5, p. 1281–1285, 2011.

MAYER, B.; KELLER, F.; SYROVETS, T.; WITTAU, M. Estimation of half-life periods in nonlinear data with fractional polynomials. **Statistical Methods in Medical Research**, v. 25, n. 5, p. 1791–1803, 2016.

MCCULLOCH, C. E.; SEARLE, S. R.; NEUHAUS, J. M. **Generalized, Linear, and Mixed Models**. 2. ed. Hoboken, New Jersey: John Wiley & Sons, 2008. 384 p.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5. ed. Hoboken, New Jersey: John Wiley & Sons, 2012. 645 p.

MUN, J.; LINDSTROM, M. J. Diagnostics for repeated measurements in linear mixed effects models. **Statistics in Medicine**, v. 32, n. 8, p. 1361–1375, 2013.

NOBRE, J. S. Métodos de diagnóstico para modelos lineares mistos. São Paulo, 2004. Dissertação (Mestrado) - IME/USP.

NOBRE, J. S. Testes de hipóteses para componentes de variância utilizando estatísticas U. São Paulo, 2007. Tese (Doutorado) - IME/USP.

NOBRE, J. S.; SINGER, J. M. Residual analysis for linear mixed models. **Biometrical Journal**, v. 49, n. 6, p. 863–875, 2007.

NOBRE, J. S.; SINGER, J. M. Leverage analysis for linear mixed models. **Journal of Applied Statistics**, v. 38, n. 5, p. 1063–1072, 2011.

NOBRE, J. S.; SINGER, J. M.; SILVAPULLE, M. J. U-tests for variance components in one-way random effects models. In: BALAKRISHNAN, E. N.; PENA, E.; SILVAPULLE, M. J. (Ed.). **Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen**. IMS Collections, 2008. p. 197–210.

OSORIO, F.; PAULA, G. A.; GALEA, M. Assessment of local influence in elliptical linear models with longitudinal structure. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4354–4368, 2007.

PACE, R. K. Nonparametric methods with applications to hedonic models. **The Journal of Real Estate Finance and Economics**, v. 7, p. 185–204, 1993.

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

PAULA, G. A. **Modelos de Regressão: Com Apoio Computacional**. São Paulo: IME-USP, 2013.

PEÑA, D. A new statistic for influence in linear regression. **Technometrics**, v. 17, n. 1, p. 1–12, 2005.

PINHEIRO, J. C.; BATES, D. M. **Mixed-Effects Models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528 p.

PINHEIRO, J. C.; BATES, D. M.; DEBROY, S.; SARKAR, D.; R CORE TEAM. **nlme: Linear and Nonlinear Mixed Effects Models**, 2018. R package version 3.1-137.

PINHEIRO, J. C.; LIU, C.; WU, Y. N. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. **Journal of Computational and Graphical Statistics**, v. 10, n. 2, p. 249–276, 2001.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018.

RATKOWSKY, D. A. **Nonlinear regression modeling: A unified practical approach**. New York: Marcel Dekker, 1983.

ROBERT-GRANIÉ, C.; MAZA, E.; RUPP, R.; FOULLEY, J. Use of fractional polynomial for modelling somatic cell scores in dairy cattle. In: **Proceedings of the 7th World Congress of Genetics Applied to Livestock Production**. Montpellier, France: CD-ROM Communication No 16-05, ISBN 9782738010520,, 2002.

ROBINSON, G. K. That BLUP is a good thing: the estimation of random effects. **Statistical Science**, v. 6, n. 1, p. 15–32, 1991.

ROCHA, F. M. M. Seleção de estruturas de covariância para dados com medidas repetidas. São Paulo, 2004. Dissertação (Mestrado) - IME/USP.

ROCHA, F. M. M.; SINGER, J. M. Selection of terms in random coefficient regression models. **Journal of Applied Statistics**, v. 45, n. 2, p. 225–242, 2018.

ROYSTON, P.; ALTMAN, D. G. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. **Applied Statistics**, v. 43, p. 429–467, 1994.

ROYSTON, P.; ALTMAN, D. G.; SAUERBREI, W. Dichotomizing continuous predictors in multiple regression: a bad idea. **Statistics in Medicine**, v. 25, n. 1, p. 127–141, 2006.

ROYSTON, P.; AMBLER, G.; SAUERBREI, W. The use of fractional polynomials to model continuous risk variables in epidemiology. **International Journal of Epidemiology**, v. 28, n. 5, p. 964–974, 1999.

ROYSTON, P.; SAUERBREI, W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. **Statistics in Medicine**, v. 23, p. 2509–2525, 2004.

ROYSTON, P.; SAUERBREI, W. Improving the robustness of fractional polynomial models by preliminary covariate transformation: A pragmatic approach. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4240–4253, 2007.

ROYSTON, P.; SAUERBREI, W. **Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables**. Chichester, England: John Wiley & Sons, 2008. 303p.

SARKAR, D. **Lattice: Multivariate Data Visualization with R**. New York: Springer, 2008. 265 p.

SAUERBREI, W.; MEIER-HIRMER, C.; BENNER, A.; ROYSTON, P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. **Computational Statistics & Data Analysis**, v. 50, p. 3464–3485, 2006.

SAUERBREI, W.; ROYSTON, P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. **Journal of the Royal Statistical Society, A**, v. 162, Part 1, p. 71–94. Corrigendum: 2002; 165, Part 2: 399–400, 1999.

SAUERBREI, W.; ROYSTON, P.; BINDER, H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. **Statistics in Medicine**, v. 26, p. 5512–5528, 2007a.

SAUERBREI, W.; ROYSTON, P.; LOOK, M. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. **Biometrical Journal**, v. 49, p. 453–473, 2007b.

SAVALLI, C.; PAULA, G. A.; CYSNEIROS, F. J. A. Assessment of variance components in elliptical linear mixed models. **Statistical Modelling**, v.6, n.1, p.59–76, 2006.

SCHÜTZENMEISTER, A.; PIEPHO, H.-P. Residual analysis of linear mixed models using a simulation approach. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1405–1416, 2012.

SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, v. 6, n. 2, p. 461–464, 1978.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance Components**. New York: John Wiley, 1992. 501p.

SELF, S. G.; LIANG, K. Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. **Journal of the American Statistical Association**, v. 82, n. 398, p. 605–610, 1987.

SHKEDY, Z.; AERTS, M.; MOLENBERGHS, G.; BEUTELS, P.; VAN DAMME, P. Modelling age-dependent force of infection from prevalence data using fractional polynomials. **Statistics in Medicine**, v. 25, n. 9, p. 1577–1591, 2006.

SINGER, J. M.; NOBRE, J. S.; ROCHA, F. M. M. **Análise de Dados Longitudinais: Versão Parcial Preliminar**. São Paulo: IME-USP, 2018.

SINGER, J. M.; ROCHA, F. M. 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.

STALEY, J. R.; BURGESS, S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. **Genetic epidemiology**, v. 41, n. 4, p. 341–352, 2017.

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

STRAM, D. O.; LEE, J. W. Corrections: Variance components testing in the longitudinal mixed effects model. **Biometrics**, v. 51, n. 3, p. 1196, 1995.

TAN, F. E. S.; OUWENS, M. J. N.; BERGER, M. P. F. Detection of influential observations in longitudinal mixed effects regression models. **Journal of the Royal Statistical Society, D**, v. 50, n. 3, p. 271–284, 2001.

TAN, M.; QU, Y.; KUTNER, M. H. Model diagnostics for marginal regression analysis of correlated binary data. **Communications in Statistics - Simulation and Computation**, v. 26, n. 2, p. 539–558, 1997.

TAN, Q.; THOMASSEN, M.; HJELMBORG, J. V. B.; CLEMMENSEN, A.; ANDERSEN, K. E.; PETERSEN, T. K.; MCGUE, M.; CHRISTENSEN, K.; KRUSE, T. A. A growth curve model with fractional polynomials for analysing incomplete time-course data in microarray gene expression studies. **Advances in bioinformatics**, v. 2011, n. ID 261514, p. 1–6, 2011.

THE EUROPEAN COLLABORATIVE STUDY. Height, weight, and growth in children born to mothers with HIV-1 infection in Europe. **Pediatrics**, v. 111, n. 1, p. e52–e60, 2003.

TILLING, K.; MACDONALD-WALLIS, C.; LAWLOR, D. A.; HUGHES, R. A.; HOWE, L. D. Modelling childhood growth using fractional polynomials and linear splines. **Annals of Nutrition and Metabolism**, v. 65, n. 2–3, p. 129–138, 2014.

VANDERSTICHEL, R.; DOHOO, I.; SANCHEZ, J.; SITHOLE, F.; KEEFE, G.; STRYHN, H. Predicting the effect of anthelmintic treatment on milk production of dairy cattle in Canada using an *Ostertagia ostertagi* ELISA from individual milk samples. **Preventive veterinary medicine**, v. 111, n. 1–2, p. 63–75, 2013.

VENEZUELA, M. K. Modelos Lineares Generalizados para Análise de dados com medidas repetidas. São Paulo, 2004. Dissertação (Mestrado) - IME/USP.

VERBEKE, G.; LESAFFRE, E. Large sample properties of the maximum likelihood estimators in linear mixed models with misspecified random-effects distributions. Rel. téc., Technical report, Biostatistical Centre for Clinical Trials, Catholic University of Leuven, Belgium, 1996.

VERBEKE, G.; MOLENBERGHS, G. **Linear Mixed Models for Longitudinal Data**. New York: Springer-Verlag, 2000. 528 p.

VONESH, E. F.; CARTER, R. L. Mixed-effects nonlinear regression for unbalanced repeated measures. **Biometrics**, v. 48, p. 1–17, 1992.

VU, H. T. V.; ZHOU, S. Generalization of likelihood ratio tests under nonstandard conditions. **The Annals of Statistics**, v. 25, n. 2, p. 897–916, 1997.

WATERNAUX, C.; LAIRD, N. M.; WARE, J. H. Methods for analysis of longitudinal data: blood-lead concentrations and cognitive development. **Journal of the American Statistical Association**, v. 84, n. 405, p. 33–41, 1989.

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

WEISBERG, S. **Applied Linear Regression**. 3. ed. Hoboken, New Jersey: John Wiley & Sons, 2005. 310 p.

WEN, X.; KLEINMAN, K.; GILLMAN, M. W.; RIFAS-SHIMAN, S. L.; TAVERAS, E. M. Childhood body mass index trajectories: modeling, characterizing, pairwise correlations and socio-demographic predictors of trajectory characteristics. **BMC medical research methodology**, v. 12, p. 1–13, 2012.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. 2. ed. New York: Springer-Verlag, 2016. 260 p.

WOLFINGER, R. Covariance structure selection in general mixed models. **Communications in Statistics-Simulation and Computation**, v. 22, n. 4, p. 1079–1106, 1993.

WONG, E. S.; WANG, B. C.; GARRISON, L. P.; ALFONSO-CRISTANCHO, R.; FLUM, D. R.; ARTERBURN, D. E.; SULLIVAN, S. D. Examining the BMI-mortality relationship using fractional polynomials. **BMC medical research methodology**, v. 11, n. 1, p. 175, 2011.

WU, C. F. J.; HAMADA, M. S. **Experiments: Planning, Analysis, and Optimization**. 2. ed. New Jersey: John Wiley & Sons, 2009. 716 p.

ZARE, K.; RASEKH, A. Diagnostic measures for linear mixed measurement error models. **SORT-Statistics and Operations Research Transactions**, v. 35, n. 2, p. 125–144, 2011.

ZEWOTIR, T. Multiple cases deletion diagnostics for linear mixed models. **Communications in Statistics - Theory and Methods**, v. 37, n. 7, p. 1071–1084, 2008.

ZHANG, Z. Multivariable fractional polynomial method for regression model. **Annals of Translational Medicine**, v. 4, n. 9, 2016.

ZHONG, X.-P.; FUNG, W.-K.; WEI, B.-C. Estimation in linear models with random effects and errors-in-variables. **Annals of the Institute of Statistical Mathematics**, v. 54, n. 3, p. 595–606, 2002.

ZHU, H.; IBRAHIM, J. G.; LEE, S.; ZHANG, H. Perturbation selection and influence measures in local influence analysis. **The Annals of Statistics**, v. 35, n. 6, p. 2565–2588, 2007.