



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE ESTATÍSTICA

NATHALIA LIMA CHAVES

**MODELOS PARA DADOS PRÉ-TESTE/PÓS-TESTE NO CONTEXTO
CLÁSSICO E BAYESIANO**

FORTALEZA
2014

NATHALIA LIMA CHAVES

MODELOS PARA DADOS PRÉ-TESTE/PÓS-TESTE NO CONTEXTO
CLÁSSICO E BAYESIANO

Monografia apresentada ao curso de Estatística do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Juvêncio Santos Nobre

FORTALEZA

2014

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca do Curso de Matemática

-
- C439m Chaves, Nathalia Lima
Modelos para dados pré-testes/pós-teste no contexto clássico e bayesiano / Nathalia Lima Chaves. – 2014.
83 f. : il., enc. ; 31 cm
- Monografia (graduação) – Universidade Federal do Ceará, Centro de Ciências, Departamento de Estatística e Matemática Aplicada, Fortaleza, 2014.
Orientação: Prof. Dr. Juvêncio Santos Nobre.
1. Modelos estatísticos. 2. Modelos lineares (Estatística). 3. Estudos longitudinais. I. Título.

CDD 519.5

NATHALIA LIMA CHAVES

MODELOS PARA DADOS PRÉ-TESTE/PÓS-TESTE NO CONTEXTO
CLÁSSICO E BAYESIANO

Monografia apresentada ao curso de Estatística do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharela em Estatística.

Aprovada em ___/___/_____.

BANCA EXAMINADORA

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

Prof. Dr. Maurício Araújo Mota
Universidade Federal do Ceará (UFC)

Profa. Dra. Sílvia Maria de Freitas
Universidade Federal do Ceará (UFC)

“Gosto dos venenos mais lentos, das bebidas mais amargas, das drogas mais poderosas, das ideias mais insanas, dos pensamentos mais complexos, dos sentimentos mais fortes...tenho um apetite voraz e os delírios mais loucos.

Você pode até me empurrar de um penhasco que eu vou dizer:

- E daí? Eu adoro voar!”

Clarice Lispector

“Depois que cansei de procurar, aprendi a encontrar. Depois que um vento me opôs resistência, velejo com todos os ventos.”

Friedrich Nietzsche

“Também é preciso ferrugem: não basta ser afiado! Senão sempre dirão: - É jovem demais!”

Friedrich Nietzsche

A Deus pela dádiva da vida

Aos meus pais pelo amor incondicional

À Roberta pela amizade e carinho indispensáveis

Aos meus mestres pelos valiosos ensinamentos.

AGRADECIMENTOS

Meus sinceros agradecimentos:

A Deus por me conceder saúde, por iluminar constantemente o meu caminho e as minhas decisões e por me presentear com pessoas tão especiais.

À minha mãe, pelo amor (“Um amor mais forte que tudo, mais obstinado que tudo, mais duradouro que tudo, é somente o amor de mãe”), carinho, pela educação, pela paciência e doçura, pela inteligência, delicadeza, e por exercer com maestria, quando necessário, a função de mãe e pai ao mesmo tempo. Mãe: Teus braços sempre se abrem quando preciso de um abraço. Teu coração sabe compreender quando preciso de uma amiga. Teus olhos sensíveis se endurecem quando preciso de uma lição. Tua força e teu amor me dirigiram pela vida e me deram as asas que precisava para voar. Tudo o que sou hoje devo aos teus ensinamentos e aos teus bons exemplos.

Ao meu pai (*In Memoriam*), pelo amor e carinho, pela educação, pela total dedicação à família e pelo exemplo de honestidade e companheirismo. Obrigada por estar tão presente na minha vida através dos teus valiosos ensinamentos e por caminhar ao meu lado, de uma forma ou de outra. Tenho certeza que te terei sempre ao meu lado, torcendo pelo meu sucesso e abençoando as minhas escolhas.

À minha irmã, pelo seu bom humor e por tornar o meu dia a dia muito mais divertido. À toda minha família, pelo exemplo de união e pelo amor dedicado.

À Roberta, pela amizade, pelo carinho, pela compreensão e parceria, pela atenção (diga-se de passagem é a pessoa mais atenciosa que conheço), pela inteligência e sensatez compartilhadas, pelos inúmeros conselhos e aprendizados, pelos momentos bons e pelos momentos difíceis em que estive sempre ao meu lado, por acreditar em mim e me dá forças pra seguir em frente, por ajudar a realizar alguns dos meus planos e por dividir comigo os meus sonhos, mesmo aqueles que parecem mais distantes. Obrigada por tudo, vencemos mais uma etapa.

À Veriana, por quem tenho uma imensa admiração, pela amizade verdadeira, pelo carinho, pelos trabalhos em grupo, ou melhor em trio, pelas horas de estudo, pelos maravilhosos e divertidos momentos vivenciados e por ter ajudado a tornar muito mais

agradável essa caminhada.

Ao professor Júlio Francisco Barros Neto, pela amizade, pela preocupação e atenção constantes, por ser um dos grandes responsáveis pelo meu amadurecimento pessoal e por ter me dado a oportunidade de fazer parte do PET-ESTATÍSTICA sob a sua tutoria.

A todos que vivenciaram comigo a intensa experiência de participar do PET-ESTATÍSTICA, em especial os meus agradecimentos ao Eriton, ao Sidarta, a Raquel, a Patrícia e ao Gerson, pela parceria e pela contribuição com o meu crescimento pessoal.

Ao professor André Jalles Monteiro, meu primeiro orientador, por me fazer enxergar a Estatística sob o ponto de vista filosófico, além das fórmulas e dos números.

Ao meu orientador Juvêncio Santos Nobre pela confiança que em mim foi depositada, pela paciência, pelas dúvidas esclarecidas prontamente, pelas conversas e pelos conselhos valiosos, pelas tentativas de me tranquilizar (mesmo estando muito tensa), pelo incentivo e principalmente por acreditar em mim.

Ao professor João Maurício Araújo Mota pelo carinho, pela paciência e pelos conhecimentos compartilhados. É uma honra ter sido aluna desse mestre que sabe ser professor de verdade, não hesita em ensinar tudo o que sabe e transmite aos seus alunos paixão pelo que faz.

Ao professor José Aílton Alencar Andrade pela ajuda e disponibilidade e por enriquecer este trabalho com os seus conhecimentos bayesianos.

A todos aqueles que fazem parte do Departamento de Estatística e Matemática Aplicada (DEMA) da Universidade Federal do Ceará. Em especial, aos professores: Sílvia Maria de Freitas, Ana Maria Souza de Araújo e Carlos Robson Bezerra de Medeiros; às secretárias: Márgeri Cavalcante França e Luísa Cristina Ribeiro Silva.

A todos, que direta ou indiretamente, contribuíram com essa conquista.

RESUMO

Em muitas situações práticas, o interesse reside em modelar dados oriundos de estudos com medidas repetidas. Para este fim, é importante o desenvolvimento de técnicas para modelar, ajustar e avaliar a qualidade do modelo utilizado para dados com estas características. Grande parte do esforço empregado na análise de dados com medidas repetidas está relacionado com a modelagem da estrutura de dependência intra-unidades experimentais. Dentro da classe de estudos com medidas repetidas, vale ressaltar os estudos com dados do tipo pré-teste/pós-teste. Modelos estatísticos para dados pré-teste/pós-teste foram considerados por diversos autores. Tais modelos são caracterizados como estudos longitudinais cujas observações são tomadas antes e após algum tipo de intervenção, um tratamento, por exemplo, para seu estudo de efeitos ao longo do tempo. São diversas as situações em que experimentos com dados pré e pós-teste são realizados e modelos específicos para este tipo de abordagem já são muito consagrados. A partir do uso de um conjunto de dados previamente analisado (Sef, 1999), dado a existência de observações influentes e alavanca no contexto de modelos lineares mistos, conforme destacado em Nobre e Singer (2007, 2011), propomos apresentar o uso dos modelos lineares mistos, no contexto clássico e bayesiano, como uma das ferramentas mais utilizadas para modelar dados do tipo pré-teste/pós-teste, com o objetivo de obter modelos mais robustos para fins de predição.

Palavras-chave: Estudos longitudinais, Medidas repetidas, Dados pré-teste e pós-teste, Modelos lineares mistos.

ABSTRACT

In many practical situations, the interest lies in modeling data from studies with repeated measures. To this end, it is important to develop techniques for shaping, setting and assess the quality of the data model used for these criteria. Much of the effort employed in data analysis with repeated measures is related to the dependence structure modeling intra-experimental units. Within the class of studies with repeated measures, it is worth mentioning the studies with data type pré-teste/pós-teste. Statistical models for data pré-teste/pós-teste been considered by several authors. These models are characterized as longitudinal studies whose observations are taken before and after any type of intervention, treatment, for example, to study their effects over time. There are different ways in which experiments with data pre-and post-test are conducted and specific models for this approach are already very devoted. From the use of two datasets previously analyzed (Sef, 1999, Singer et al., 1988), given the existence of influential observations and leverage in the context of linear mixed models, as highlighted in Noble and Singer (2007, 2011), propose to introduce the use of mixed linear models, Bayesian and the classical context, as one of the most used tools to model data type pré-teste/pós-teste, in order to obtain more robust models for prediction.

Keywords: Longitudinal studies, Repeated measures, Experimental units, Pre-test and post-test data, Linear mixed models, Flexible bayesian methodology.

LISTA DE FIGURAS

1	Modelos de escovas do tipo monobloco utilizadas no estudo.	17
2	Gráfico de dispersão múltiplo entre os índices de placa bacteriana pré e pós-tratamento para as escovas convencional e monobloco.	19
3	Boxplots dos índices de placa bacteriana pré/pós-tratamento para cada sessão.	19
4	Relação entre vendas e descontos.	25
5	Ajuste do modelo final	63
6	(a) Resíduo marginal (b) EBLUP	64
7	Resíduo condicional padronizado e envelope simulado com 95% para o resíduo com confundimento mínimo.	64
8	(a)Distância de Mahalanobis (b)Valores preditos para cada tipo de escova .	65
9	Distribuições a priori	68
10	Distribuições a posteriori dos parâmetros com priori Gama(0,01; 0,01) . . .	69
11	Distribuições a posteriori dos parâmetros com priori Exponencial(0,001) . .	70
12	Trajetórias completas da simulação dos parâmetros com priori Gama(0,01; 0,01)	71
13	Trajetórias completas da simulação dos parâmetros com priori Exponencial(0,001)	72

LISTA DE TABELAS

1	Disposição esquemática dos dados em estudos com medidas repetidas. . . .	15
2	Índice de placa bacteriana dentária.	18
3	Disposição esquemática dos dados em estudos com medidas repetidas. . . .	22
4	Uso dos diferentes resíduos na análise de diagnóstico.	43
5	Estimativas dos parâmetros e intervalos de confiança a um nível de confiança de 95% ao eliminar as unidades experimentais #12 e #29.	65
6	Estimativas, média e mediana a posteriori e intervalos de credibilidade de 95% do modelo final.	68

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Análise de medidas repetidas	14
1.1.1	Experimentos com dados do tipo pré e pós-teste	16
2	ANÁLISE DE MEDIDAS REPETIDAS	21
2.1	Introdução	21
2.2	Experimentos do tipo pré-teste e pós-teste	22
3	MODELOS LINEARES MISTOS	24
3.1	Introdução e Motivação	24
3.2	Especificações do modelo	26
3.3	Estruturas de covariância	27
3.4	Inferência Estatística	31
3.4.1	Estimação por máxima verossimilhança	31
3.4.2	BLUE e BLUP	33
3.4.3	Teste de hipóteses e Critérios de informação	34
4	ANÁLISE DE DIAGNÓSTICO	37
4.1	Modelo normal linear	37
4.1.1	Análise de Resíduos	37
4.1.2	Análise de Sensibilidade	39
4.1.2.1	Influência	39
4.1.2.2	Influência Local	40

4.2	Modelos Lineares Mistos	41
4.2.1	Análise de Resíduos	41
4.2.2	Análise de Sensibilidade	44
4.2.2.1	Eliminação de observações	44
4.2.2.2	Pontos de Alavanca	46
5	ABORDAGEM BAYESIANA	49
5.1	Introdução	49
5.2	Teorema de Bayes	50
5.3	Informações a priori	51
5.3.1	Distribuições a priori subjetivas	51
5.3.2	Distribuições a priori conjugadas	51
5.3.3	Distribuições a priori não-informativas	52
5.3.3.1	Priori Uniforme	53
5.3.3.2	Priori de Jeffreys	53
5.4	Metodologia inferencial	54
5.4.1	Predição pontual	54
5.4.2	Funções de perda	54
5.4.3	Predição Intervalar	55
5.5	Abordagem bayesiana para modelos lineares mistos	56
5.5.1	Relação entre o BLUE e o BLUP e os estimadores bayesianos	58
6	APLICAÇÃO	60
6.1	Índice de placa bacteriana	60
7	CONCLUSÃO	73
	REFERÊNCIAS	74

1 INTRODUÇÃO

1.1 Análise de medidas repetidas

Os experimentos com medidas repetidas distinguem-se dos demais experimentos pelo fato de que uma sequência de medidas de uma ou mais variáveis respostas podem ser obtidas em uma mesma unidade experimental em diferentes condições de avaliação, constituindo um perfil dessa unidade. A importância do estudo com medidas repetidas reside na modelagem da possível correlação existente entre as observações intra unidades experimentais. Tais experimentos diferenciam-se dos estudos do tipo transversal (que envolvem uma única observação, realizada num instante especificado, da variável resposta por unidade experimental de cada população de interesse), pois permitem a realização de análises dessa natureza necessitando de um número menor de unidades experimentais e permitindo a avaliação de mudanças que ocorrem intra e entre essas unidades experimentais.

Os estudos com medidas repetidas tem como principais vantagens os seguintes fatos: requerem menos unidades experimentais do que planejamentos completamente casualizados, proporcionam condições mais adequadas para o estudo de covariáveis que possam ter influência na variável resposta, melhoram, em geral, a precisão de contrastes associados às diferenças entre os valores médios da variável resposta de diferentes tratamentos e permitem o estudo da mudança do comportamento da resposta média da unidade experimental nos diferentes tratamentos (incorpora informação sobre a variação individual na análise).

De acordo com Rocha (2004), as principais desvantagens de estudos com medidas repetidas são: a primeira é que a análise é mais complicada por requerer que as correlações das medidas realizadas no mesmo indivíduo sejam modeladas. A segunda desvantagem está relacionada com a possível presença de dados incompletos ou omissos (“missing data”), o que na prática ocorre com frequência em estudos dessa natureza.

Estudos com medidas repetidas são classificados como regulares em relação às condições

de avaliação quando o intervalo entre duas medidas consecutivas é constante; em caso contrário, são chamados de irregulares. Quanto ao planejamento, eles podem ser classificados como balanceados em relação às condições de avaliação se as medidas forem obtidas nos mesmos instantes de avaliação em todas as unidades experimentais; em caso contrário, são ditos não balanceados. Em relação às condições de avaliação, esses estudos são completos se não houver observações perdidas (dados omissos) e incompletos, se existirem. Estudos com dados incompletos ocorrem quando, por alguma razão, as observações de uma ou mais unidades de investigação, não são obtidas sob alguma condição de avaliação.

Dados oriundos de estudos com essas características são apresentados através de vetores p -variados $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})^\top$, $i = 1, \dots, K$ e $j = 1, \dots, n_i$, em que y_{ijk} representa a variável resposta observada da j -ésima unidade (indivíduo) do i -ésimo grupo na k -ésima ocasião (dose, tempo, etc.). Na Tabela 1 apresenta-se, como ilustração, uma disposição esquemática dos dados de estudos balanceados e completos com medidas repetidas.

Tabela 1: Disposição esquemática dos dados em estudos com medidas repetidas.

Grupos	Unidade Experimental	Condições de avaliação			
		1	2	...	p
1	1	y_{111}	y_{112}	...	y_{11p}
1	2	y_{121}	y_{122}	...	y_{12p}
⋮	⋮	⋮	⋮	...	⋮
1	n_1	y_{1n_11}	y_{1n_12}	...	y_{1n_1p}
2	1	y_{211}	y_{212}	...	y_{21p}
2	2	y_{221}	y_{222}	...	y_{22p}
⋮	⋮	⋮	⋮	...	⋮
2	n_2	y_{2n_21}	y_{2n_22}	...	y_{2n_2p}
⋮	⋮	⋮	⋮	⋮	⋮
K	1	y_{K11}	y_{K12}	...	y_{K1p}
K	2	y_{K21}	y_{K22}	...	y_{K2p}
⋮	⋮	⋮	⋮	...	⋮
K	n_K	y_{Kn_K1}	y_{Kn_K2}	...	y_{Kn_Kp}

Dentre os diversos métodos e modelos que tem sido propostos na literatura para tratar dados dessa natureza, os modelos lineares e não-lineares mistos são os mais utilizados, como pode ser visto em Lindstrom e Bates (1990), Searle et al. (1992), Longford (1993), Jones (1993), Vonesh e Chinchilli (1996), Verbeke e Molenberghs (2000), Pinheiro e Bates (2000), Cox e Solomon (2003), dentre outros. Entretanto, segundo Wu e Zhang (2006), em algumas situações os modelos paramétricos podem ser restritivos ou limitados. Assim, para superar essa dificuldade, técnicas não-paramétricas para a análise de medidas repetidas também tem sido desenvolvidas, vide por exemplo, em Härdle (1990), Wahba

(1990), Green e Silverman (1994), Simonoff (1996), Eubank (1988, 1999), Efromovich (1999), Härdle et al. (2000), Ruppert et al. (2003).

1.1.1 Experimentos com dados do tipo pré e pós-teste

Diversos planejamentos com medidas repetidas são corriqueiros, no entanto, o mais usual são aqueles em que pesquisadores tomam medidas repetidas em diferentes tempos em relação à mesma unidade experimental, tendo o objetivo de estudar o seu comportamento no decorrer do tempo, caracterizando as medidas repetidas no tempo (ou planejamentos longitudinais). Dentre os estudos longitudinais, vale ressaltar o estudo com dados do tipo pré-teste/pós-teste que é definido como sendo uma abordagem experimental em que uma medida é obtida antes, medida pré-teste, e após uma intervenção (um tratamento, por exemplo), medida pós-teste, cujo objetivo é verificar o efeito de tal intervenção ao longo do tempo.

Segundo Bonate (2000), há em geral dois tipos de experimentos envolvendo dados de natureza pré-teste e pós-teste. O primeiro tipo ocorre quando um indivíduo é mensurado em duas ocasiões diferentes e o pesquisador tem interesse em avaliar se há diferença entre a primeira e a segunda mensuração. Esse tipo de estudo é dito ser não-controlado, porque não há grupo de controle a ser usado para comparação. Se houver diferença entre as medidas pré-teste e pós-teste, é difícil determinar se tal alteração foi devido ao uso de instrumento inapropriado para a medição ou devido a uma alteração real do indivíduo em estudo. O segundo tipo ocorre quando os indivíduos são submetidos a uma intervenção qualquer antes da medição do pós-teste, após a conclusão do pré-teste. Uma situação que exemplifica o segundo tipo de experimento é a seguinte: Em um ensaio clínico, pacientes esquizofrênicos são escolhidos aleatoriamente para receber um dentre dois tratamentos, haloperidol ou um novo medicamento. A princípio, a gravidade de cada paciente é avaliada e, em seguida, cada indivíduo recebe um dos dois medicamentos durante 6 semanas. No final do estudo, cada indivíduo é avaliado quanto a gravidade da sua doença. Neste caso, a intervenção foi o medicamento utilizado. Este é um estudo controlado, em que o tratamento de interesse é comparado com um tratamento de efeitos conhecidos. Note que o grupo controle pode receber o tratamento padrão ou placebo.

A seguir, apresentaremos o exemplo que será utilizado ao longo desse trabalho.

Exemplo 1 (Índice de placa bacteriana): Um estudo longitudinal foi realizado na Faculdade de Odontologia da Universidade de São Paulo, para comparar dois tipos de escova de dentes, monobloco e convencional, (Parizzoto, 1999) com relação à eficácia na remoção de placa bacteriana. Com esta finalidade, foram observadas 32 crianças em 4 sessões quinzenais, uma das quais correspondente à avaliação inicial. As crianças foram alocadas em dois grupos de tamanhos iguais, cada um submetido ao tratamento com uma das escovas. Durante o período de observação, cada criança utilizou a mesma escova que lhe foi dada na primeira sessão. Em cada sessão de avaliação, mediu-se o índice de placa bacteriana antes (pré-tratamento) e depois (pós-tratamento) da escovação. Os dados referentes ao estudo encontram-se na Tabela 2 e mais detalhes podem ser encontrados em Sef (1999) e Nobre (2004). Na Figura 1 apresenta-se as escovas do tipo monobloco, que foram propostas para o estudo em questão. Na Figura 2 apresenta-se os gráficos de dispersão entre os índices de placa bacteriana pré-tratamento (x) e pós-tratamento (y) para os dois tipos de escovas em questão. Na Figura 3, os boxplots dos índices de placa bacteriana pré/pós-tratamento para cada tipo de escova indicam a heterogeneidade dos dados, dando indícios da existência de observações discrepantes.

Figura 1: Modelos de escovas do tipo monobloco utilizadas no estudo.



Fonte: <http://www.redetec.org.br/inventabrasil/escomon.htm>

Tabela 2: Índice de placa bacteriana dentária.

Criança	Escova	1ª Sessão		2ª Sessão		3ª Sessão		4ª Sessão	
		Antes	Depois	Antes	Depois	Antes	Depois	Antes	Depois
1	Convencional	1,05	1,00	1,13	0,84	1,15	0,86	1,13	0,94
2	Convencional	1,07	0,62	0,92	0,62	1,02	0,57	1,15	0,85
3	Convencional	0,82	0,62	1,52	1,07	1,39	0,97	1,78	1,39
4	Convencional	1,37	0,90	1,65	1,20	1,75	1,40	1,92	1,67
5	Convencional	1,97	1,52	1,30	1,07	1,50	1,15	1,65	1,37
6	Convencional	1,30	0,82	1,17	0,70	0,75	0,50	1,47	1,12
7	Convencional	1,61	1,19	1,52	1,13	1,22	1,00	1,63	1,22
8	Convencional	1,02	0,73	1,08	0,64	0,94	0,73	1,14	0,97
9	Convencional	1,62	1,25	1,45	1,10	1,10	0,75	1,70	1,32
10	Convencional	1,65	1,22	1,57	1,22	1,47	1,10	1,62	1,17
11	Convencional	1,02	0,78	0,60	0,47	0,88	0,75	1,36	1,08
12	Convencional	0,71	0,60	1,13	0,39	0,84	0,65	1,65	1,31
13	Convencional	1,70	1,55	1,85	1,37	1,87	1,55	1,60	1,30
14	Convencional	1,30	1,02	1,65	0,97	1,72	1,20	1,37	1,22
15	Convencional	1,40	0,80	1,83	1,03	1,76	1,38	1,96	1,15
16	Convencional	1,40	1,12	1,25	0,67	1,50	1,10	1,50	1,22
17	Monobloco	1,66	1,63	1,36	1,16	1,52	0,88	1,41	1,20
18	Monobloco	1,02	0,80	0,92	0,82	1,10	0,76	1,28	1,15
19	Monobloco	0,75	0,67	1,00	0,92	1,00	0,87	1,15	1,10
20	Monobloco	1,29	1,23	0,91	0,76	1,14	0,94	1,35	0,97
21	Monobloco	1,27	1,20	1,20	0,95	1,10	1,00	1,37	1,17
22	Monobloco	1,07	0,85	1,39	1,25	1,39	1,25	1,28	1,21
23	Monobloco	1,35	1,21	1,42	1,17	1,42	1,19	1,42	1,23
24	Monobloco	1,32	1,02	1,60	1,40	1,35	1,02	1,50	1,25
25	Monobloco	1,66	1,61	1,50	1,36	1,72	1,41	1,69	1,44
26	Monobloco	1,30	1,07	0,84	0,61	0,88	0,61	0,96	0,57
27	Monobloco	1,57	1,20	1,50	1,07	1,15	1,00	1,25	1,05
28	Monobloco	1,67	1,50	1,47	1,32	1,07	0,97	1,50	1,37
29	Monobloco	0,91	0,67	0,96	0,62	1,09	0,53	1,12	0,37
30	Monobloco	1,06	0,70	1,00	0,85	1,15	0,93	1,12	1,00
31	Monobloco	2,30	2,00	1,37	1,25	1,40	1,32	2,15	1,90
32	Monobloco	1,15	1,00	1,23	1,11	1,15	1,07	1,26	1,00

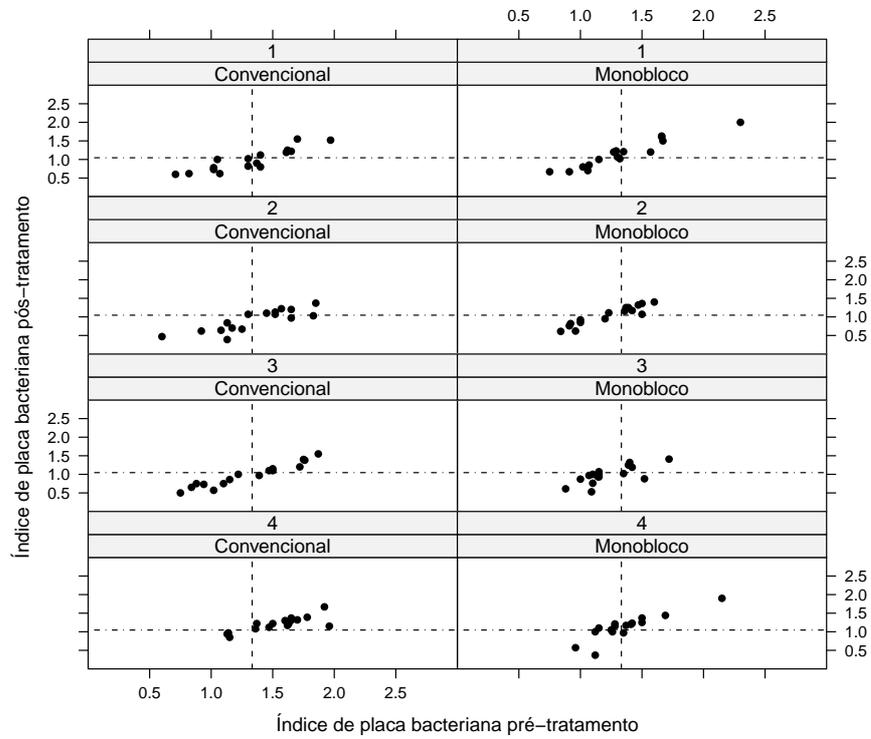


Figura 2: Gráfico de dispersão múltiplo entre os índices de placa bacteriana pré e pós-tratamento para as escovas convencional e monobloco.

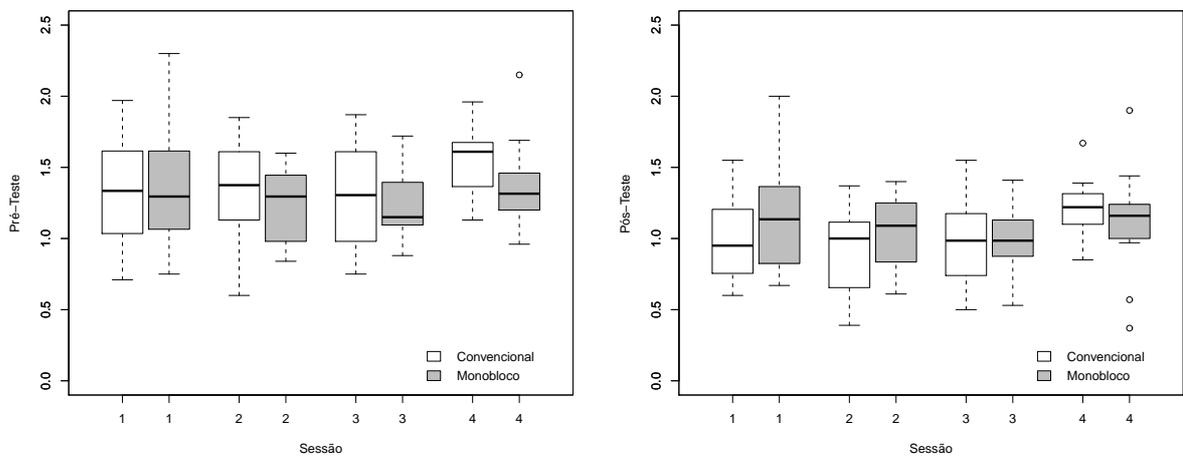


Figura 3: Boxplots dos índices de placa bacteriana pré/pós-tratamento para cada sessão.

A análise estatística de dados dessa natureza é feita utilizando-se técnicas multi ou univariadas. Como exemplos clássicos dessas técnicas de análise, vale ressaltar a análise de perfis por meio de um modelo univariado e a análise de curvas de crescimento, através de modelos lineares ou não lineares mistos, que permitem o uso de diversos tipos de estrutura para as matrizes de variância-covariância, de forma a delinear o comportamento dos perfis médios através de curvas. Dentre uma extensa gama de procedimentos usados para abordar o estudo com medidas repetidas, podemos citar o uso das seguintes técnicas: modelos lineares mistos (Crowder e Hand, 1990; Jones, 1993; Crowder e Hand, 1996; Verbeke e Molenberghs, 1997; Vonesh e Chinchilli, 1997; Singer e Andrade, 2000; Diggle et al., 2002; Davis, 2002; Demidenko, 2004), modelos lineares mistos elípticos (Pinheiro et al., 2001; Savalli et al., 2006; Osorio et al., 2007), modelos lineares mistos assimétricos (Azzalini e Dalla Valle, 1996; Azzalini e Capitanio, 2003; Sahu et al., 2003) e equações de estimação (Venezuela, 2003; Verbeke e Molenberghs, 2005; Venezuela et al., 2007; Venezuela et al., 2010). O presente trabalho irá focar na utilização do modelos lineares mistos, visto que tais modelos permitem a modelagem da matriz de covariâncias.

2 ANÁLISE DE MEDIDAS REPETIDAS

2.1 Introdução

Os experimentos com medidas repetidas distinguem-se dos demais experimentos pelo fato de que uma sequência de medidas de uma ou mais variáveis respostas pode ser obtida em uma mesma unidade experimental em diferentes condições de avaliação, constituindo um perfil da unidade. A importância do estudo com medidas repetidas reside na modelagem da possível correlação existente entre as observações intra unidades experimentais.

De acordo com Rocha (2004), as principais desvantagens de estudos com medidas repetidas são: a primeira é que a análise é mais complicada por requerer que as correlações das medidas realizadas no mesmo indivíduo sejam modeladas. A segunda desvantagem está relacionada com a possível presença de dados incompletos ou omissos (“missing data”), o que na prática ocorre com frequência em estudos dessa natureza.

Estudos com medidas repetidas são classificados como regulares em relação às condições de avaliação quando o intervalo entre duas medidas consecutivas é constante; em caso contrário, são chamados de irregulares. Quanto ao planejamento, eles podem ser classificados como balanceados em relação às condições de avaliação se as medidas forem obtidas nos mesmos instantes de avaliação em todas as unidades experimentais; em caso contrário, são ditos não balanceados. Em relação às condições de avaliação, esses estudos são completos se não houver observações perdidas (dados omissos) e incompletos, se existirem. Estudos com dados incompletos ocorrem quando, por alguma razão, as observações de uma ou mais unidades de investigação, não são obtidas sob alguma condição de avaliação.

Dados oriundos de estudos com essas características são apresentados através de vetores p -variados $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})^\top$, $i = 1, \dots, K$ e $j = 1, \dots, n_i$, onde y_{ijk} é a resposta observada da j -ésima unidade (indivíduo) do i -ésimo grupo na k -ésima ocasião

(dose,tempo,etc.).

Tabela 3: Disposição esquemática dos dados em estudos com medidas repetidas.

Grupos	Unidade Experimental	Condições de avaliação			
		1	2	...	p
1	1	y_{111}	y_{112}	...	y_{11p}
1	2	y_{121}	y_{122}	...	y_{12p}
\vdots	\vdots	\vdots	\vdots	...	\vdots
1	n_1	y_{1n_11}	y_{1n_12}	...	y_{1n_1p}
2	1	y_{211}	y_{212}	...	y_{21p}
2	2	y_{221}	y_{222}	...	y_{22p}
\vdots	\vdots	\vdots	\vdots	...	\vdots
2	n_2	y_{2n_21}	y_{2n_22}	...	y_{2n_2p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
K	1	y_{K11}	y_{K12}	...	y_{K1p}
K	2	y_{K21}	y_{K22}	...	y_{K2p}
\vdots	\vdots	\vdots	\vdots	...	\vdots
K	n_K	y_{Kn_K1}	y_{Kn_K2}	...	y_{Kn_Kp}

Dentre os diversos métodos e modelos que tem sido propostos na literatura para tratar dados dessa natureza, os modelos lineares e não-lineares mistos são os mais utilizados e embasam trabalhos desenvolvidos por Lindstrom e Bates (1990), Vonesh e Chinchilli (1996), Verbeke e Molenberghs (2000), Longford (1993), Jones (1993), Pinheiro e Bates (2000), Searle et al. (1992), Cox e Solomon (2003), dentre outros. Entretanto, segundo Wu e Zhang (2006), em algumas situações os modelos paramétricos podem ser restritivos ou limitados. Assim, para superar essa dificuldade, técnicas não-paramétricas para a análise de medidas repetidas tem sido desenvolvidas, como pode ser visto em Wahba (1990), Green e Silverman (1994), Eubank (1988, 1999), Härdle (1990), Simonoff (1996), Efromovich (1999), Härdle et al. (2000), Ruppert et al. (2003), dentre outros autores.

2.2 Experimentos do tipo pré-teste e pós-teste

Um experimento com dados pré-teste/pós-teste é definido como sendo uma abordagem experimental em que uma medida é obtida antes, medida pré-teste, e após uma intervenção (um tratamento, por exemplo), medida pós-teste. Tais experimentos tratam-se de estudos longitudinais com duas ou mais unidades de tempo distintas, cujo objetivo

é verificar o efeito de tal intervenção ao longo do tempo.

Segundo Bonate (2000), há em geral dois tipos de experimentos envolvendo dados de natureza pré-teste e pós-teste. O primeiro tipo ocorre quando um indivíduo é mensurado em duas ocasiões diferentes e o pesquisador tem interesse em avaliar se há diferença entre a primeira e a segunda mensuração. Esse tipo de estudo é dito ser não-controlado, porque não há grupo de controle a ser usado para comparação. Se houver diferença entre as medidas pré-teste e pós-teste, é difícil determinar se tal alteração foi devido ao uso de instrumento inapropriado para a medição ou devido a uma alteração real do indivíduo em estudo. O segundo tipo ocorre quando os indivíduos são submetidos a uma intervenção qualquer antes da medição do pós-teste, após a conclusão do pré-teste. Uma situação que exemplifica o segundo tipo de experimento é a seguinte: Em um ensaio clínico, pacientes esquizofrênicos são escolhidos aleatoriamente para receber um dentre dois tratamentos, haloperidol ou um novo medicamento. A princípio, a gravidade de cada paciente é avaliada e, em seguida, cada indivíduo recebe um dos dois medicamentos durante 6 semanas. No final do estudo, cada indivíduo é avaliado quanto a gravidade da sua doença. Neste caso, a intervenção foi o medicamento utilizado. Este é um estudo controlado, em que o tratamento de interesse é comparado com um tratamento de efeitos conhecidos. Note que o grupo controle pode receber o tratamento padrão ou placebo.

Os dois exemplos apresentados anteriormente são caracterizados como longitudinais do tipo pré-teste/pós-teste. O estudo realizado pela Faculdade de Odontologia da USP para comparar a eficiência na remoção de placa bacteriana de dois tipos de escovas, é caracterizado como um estudo longitudinal do tipo pré-teste/pós-teste pois é feito um acompanhamento das mesmas unidades experimentais (crianças), antes e depois do uso de umas das escovas em questão, ao longo das quatro sessões. No estudo realizado pelo Instituto de Ciências Biomédicas da Universidade de São Paulo para avaliar o efeito de $MgSO_4$ na pressão arterial média (PAM) medida em $mmHg$ de cães, tem-se uma intervenção (aplicação de $MgSO_4$) e uma população com dois estratos, que correspondem aos dois tratamentos (indometacina e nifedipina). A seguir, apresentaremos o uso dos modelos lineares mistos como forma mais usual de abordar e modelar experimentos dessa natureza.

3 *MODELOS LINEARES MISTOS*

3.1 Introdução e Motivação

Com o objetivo de ilustrar a importância do estudo dos modelos lineares mistos, será apresentado a seguir um exemplo fictício baseado na ideia apresentada em Demidenko (2004). Vamos considerar a relação entre desconto (x) e venda (y). Seja $(x_i, y_i), i = 1, \dots, K$, uma amostra coletada de vários produtos, em que foram observados seus respectivos descontos concedidos (%) e número de vendas (em milhares de unidades). Na Figura 4 (a) apresenta-se o gráfico de dispersão entre as variáveis, com o respectivo ajuste, via mínimos quadrados ordinários, de um modelo linear padrão, enquanto que na Figura 4 (b) apresenta-se o mesmo gráfico e as retas ajustadas (via mesmo modelo) para cada tipo de mercadoria, evidenciando assim um paradoxo. Tal paradoxo ocorre porque se desconsiderarmos a estrutura de unidades experimentais chegaremos a conclusão de que as vendas tendem a decrescer com a diminuição dos descontos e se considerarmos tal estrutura, as vendas tendem a aumentar com a diminuição dos descontos, o que nos parece ser mais razoável.

Via modelo de regressão usual, temos o seguinte modelo:

$$y_i = \alpha + \beta x_i + e_i, i = 1, \dots, K, \quad (3.1)$$

em que e_1, \dots, e_K representa uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância σ^2 . Em outras palavras, assume-se que as mercadorias são homogêneas. A partir da Figura 4 (b), pode-se perceber que as mercadorias não são homogêneas e variam substancialmente em termos de desconto e venda. Um modelo aparentemente mais adequado para traduzir a situação em questão é o seguinte:

$$y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}, i = 1, \dots, K; j = 1, \dots, n_i, \quad (3.2)$$

em que $y_{ij}(x_{ij})$ representam a venda (em milhares de unidades) e o desconto concedido (%), respectivamente, da j -ésima observação da i -ésima mercadoria, $\alpha_i(\beta_i)$ são os interceptos e os efeitos aleatórios associados a i -ésima mercadoria, e_{ij} é a fonte de variação associada à j -ésima observação da i -ésima mercadoria e n_i representa o número de observações da i -ésima mercadoria.

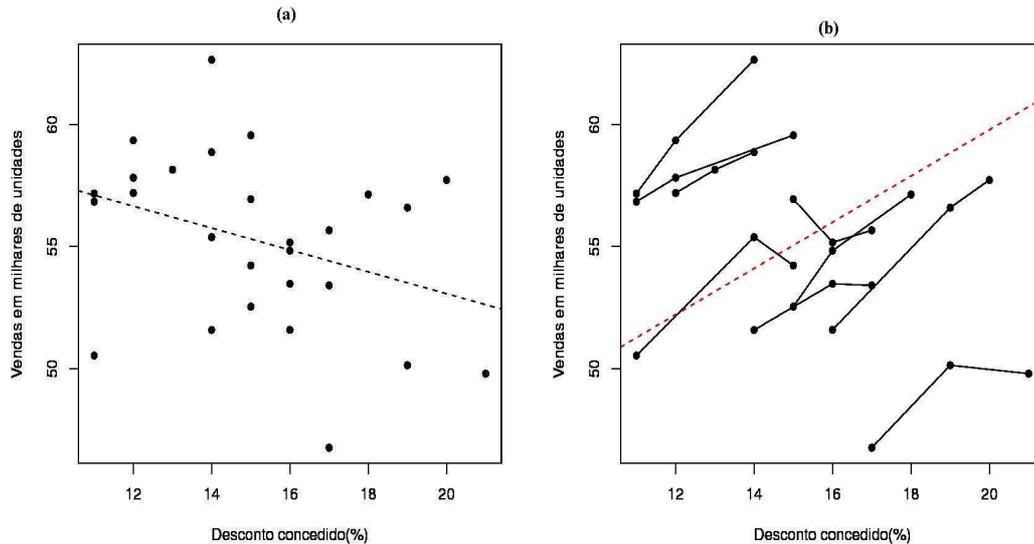


Figura 4: Relação entre vendas e descontos.

Em síntese, desconsiderar a estrutura de unidades experimentais pode levar a conclusões erradas, logo se faz necessário a utilização de modelos que permitam modelar as duas fontes de variação, intra e entre as unidades experimentais.

De uma forma geral, segundo Demidenko (2004), os modelos mistos podem ser usados, por exemplo, para os seguintes fins:

- Para modelar dados longitudinais ou estruturas de agrupamento complexas;
- Para modelar dados com múltiplas fontes de variação;
- Para modelar variedade biológica e heterogeneidade;
- Para lidar com a multidimensionalidade de parâmetros;
- Para modelar formas e imagens.

Segundo McCulloch et al. (2008), dentro do contexto dos modelos lineares mistos, alguns conceitos são de extrema importância.

- **Fatores:** São classificações que identificam a “origem” de cada observação (Sexo, Idade, Nível de exposição ao sol, por exemplo). Os fatores são classificados, de acordo com os seus níveis, de duas formas:

[i] **Fatores de Efeitos Fixos:** Os níveis do fator são fixados (escolhidos) pelo pesquisador.

[ii] **Fatores de Efeitos Aleatórios:** Os níveis do fator é uma amostra aleatória dos possíveis níveis.

- **Níveis:** São as classes individuais de cada fator (Sexo: Masculino e Feminino, por exemplo).

O interesse em classificar os indivíduos em termos de fatores e níveis é verificar o quanto cada nível afeta a variável de interesse.

3.2 Especificações do modelo

O modelo linear misto mais frequentemente encontrado na literatura apresenta a seguinte forma funcional:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, i = 1, \dots, K, \quad (3.3)$$

em que \mathbf{y}_i representa um vetor ($n_i \times 1$) de respostas da i -ésima unidade experimental, $\boldsymbol{\beta}$ é um vetor ($p \times 1$) de parâmetros (efeitos fixos), \mathbf{X}_i é uma matriz ($n_i \times p$) de especificação dos efeitos fixos, \mathbf{b}_i é um vetor ($q \times 1$) de variáveis latentes, comumente denominadas efeitos aleatórios, que refletem o comportamento individual da i -ésima unidade experimental, \mathbf{Z}_i é uma matriz ($n_i \times q$) de especificação dos efeitos aleatórios e \mathbf{e}_i é um vetor ($n_i \times 1$) de componentes aleatórios. Tipicamente assume-se que os efeitos aleatórios \mathbf{b}_i e os erros condicionais \mathbf{e}_i são independentes com

$$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D}); \mathbf{e}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \boldsymbol{\psi}_i), \quad (3.4)$$

em que $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$ e $\boldsymbol{\psi}_i = \boldsymbol{\psi}_i(\boldsymbol{\theta})$ são matrizes de variância-covariância, associadas, respectivamente, com a variabilidade entre e intra unidades experimentais.

A partir do modelo (2.3) obtém-se:

$$\mathbb{E}(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta} \quad (3.5)$$

$$\mathbb{V}(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\psi}_i \quad (3.6)$$

Este modelo também pode ser interpretado como um **modelo linear em dois estágios** como mostram Laird e Ware (1982). No primeiro estágio consideramos fixos os efeitos aleatórios \mathbf{b}_i , de forma que:

$$\mathbf{y}_i | \mathbf{b}_i \stackrel{\text{ind}}{\sim} \mathcal{N}_{n_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \boldsymbol{\psi}_i), \quad (3.7)$$

enquanto que no segundo estágio assumimos que os vetores \mathbf{b}_i são independentes com distribuição $\mathcal{N}_q(\mathbf{0}, \mathbf{D})$ e conseqüentemente, o modelo marginal é dado por:

$$\mathbf{y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \boldsymbol{\psi}_i). \quad (3.8)$$

Quando $\boldsymbol{\psi}_i = \sigma^2 \mathbf{I}_{n_i}$, com \mathbf{I}_j denotando uma matriz identidade j -dimensional, o modelo é chamado de modelo de **independência condicional homoscedástico**, indicando que as n_i observações associadas à i -ésima unidade experimental são condicionalmente independentes a \mathbf{b}_i .

O modelo (2.3) pode ser escrito de forma mais compacta através da notação matricial:

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b} + \mathbf{e}, \quad (3.9)$$

em que $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_K^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_K^\top)^\top$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_K) = \bigoplus_{i=1}^K \mathbf{Z}_i$, com \bigoplus representando o operador de soma direta (Searle, 1982), $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_K^\top)^\top$, $\mathbf{e} = (\mathbf{e}_1^\top, \dots, \mathbf{e}_K^\top)^\top$.

3.3 Estruturas de covariância

O fato de uma mesma variável ser observada ao longo de p ocasiões num mesmo indivíduo resulta na existência de correlação entre as respostas y_{ijk} , assim modelos para a análise de estudos com medidas repetidas devem apresentar estruturas de covariância que incorporem essas correlações. No contexto dos modelos lineares mistos, permite-se a consideração de formas especiais para a matriz de covariância, que buscam representar a variabilidade dos dados da forma mais real possível, ou seja, levam em consideração o fato dos dados serem independentes, dependentes, correlacionados, etc.

Segundo Diggle (1988) e Diggle et al. (2002) a matriz de covariâncias deve ser suficientemente flexível para incluir no mínimo três fontes diferentes de variação aleatória, sendo

elas: **(a)** a variação devida a efeitos aleatórios, quando as unidades de investigação formam uma amostra aleatória da população com a característica de interesse; **(b)** a variação que pode ser explicada por correlação serial, em que se esperam observações próximas mais fortemente correlacionadas que observações mais distantes e **(c)** a variação devida a erros de medição. A seguir, serão apresentadas algumas estruturas de covariâncias descritas na literatura, por intermédio de exemplos com $n_i = 4$, mais detalhes sobre os processos estocásticos que geram essas estruturas podem ser encontrados em Rocha (2004), por exemplo.

1. Estrutura Uniforme

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \tau \\ \tau & \sigma^2 + \tau & \tau & \tau \\ \tau & \tau & \sigma^2 + \tau & \tau \\ \tau & \tau & \tau & \sigma^2 + \tau \end{bmatrix}, \tau > 0.$$

2. Estrutura AR(1)

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}.$$

3. Estrutura ARMA(1,1)

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\phi & \gamma\phi^2 \\ \gamma & 1 & \gamma & \gamma\phi \\ \gamma\phi & \gamma & 1 & \gamma \\ \gamma\phi^2 & \gamma\phi & \gamma & 1 \end{bmatrix}.$$

4. Estrutura Ante-Dependência de ordem 1

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 \end{bmatrix}.$$

5. Estrutura Toeplitz

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}.$$

6. Estrutura Espacial ou Markov

$$\mathbf{V}_i(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}.$$

7. Estrutura Uniforme Heterogênea

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}.$$

8. ARH(1)

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\phi & \sigma_1\sigma_3\phi^2 & \sigma_1\sigma_4\phi^3 \\ \sigma_2\sigma_1\phi & \sigma_2^2 & \sigma_2\sigma_3\phi & \sigma_2\sigma_4\phi^2 \\ \sigma_3\sigma_1\phi^2 & \sigma_3\sigma_2\phi & \sigma_3^2 & \sigma_3\sigma_4\phi \\ \sigma_4\sigma_1\phi^3 & \sigma_4\sigma_2\phi^2 & \sigma_4\sigma_3\phi & \sigma_4^2 \end{bmatrix}.$$

9. Estrutura Toeplitz Heterogênea

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}.$$

10. Estrutura baseada em efeitos aleatórios

Para um modelo linear com coeficientes lineares e angulares aleatórios, (2.6) se reduz a

$$\mathbf{V}_i = \mathbf{Z}_i \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \mathbf{Z}_i^\top + \boldsymbol{\psi}_i,$$

com

$$\mathbf{Z}_i^\top = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \end{bmatrix}.$$

11. Não Estruturada (NE)

Neste caso nenhuma restrição é imposta aos parâmetros. Para n_i condições de avaliação, o número de parâmetros é $n_i(n_i + 1)/2$.

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}.$$

12. Componentes de Variância (VC)

$$\mathbf{V}_i(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}.$$

3.4 Inferência Estatística

3.4.1 Estimação por máxima verossimilhança

Vários métodos de estimação dos parâmetros do modelo (2.3) estão disponíveis na literatura, dentre eles convém destacar os métodos de Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR), proposto por Patterson e Thompson (1971), Harville (1977b), Robinson (1991), Searle et al. (1992) e Jiang (1996), o método de Mínimos Quadrados (MQ) e o método bayesiano detalhado em Tountenburg (1982), Maritz e Lwin (1989) e Searle et al. (1992). Supondo que \mathbf{D} e $\boldsymbol{\psi}_i$ são conhecidas todos esses métodos são equivalentes desde que no método bayesiano seja atribuída uma distribuição a priori não informativa para \mathbf{b} (Robinson, 1991). Uma abordagem bayesiana para os modelos lineares mistos será apresentada em detalhes posteriormente.

Como é conhecido, o método de máxima verossimilhança tem como base o comportamento probabilístico da fonte de variação associada ao modelo no caso dos modelos lineares usuais ou no comportamento da variável de interesse nos modelos lineares generalizados. Sendo assim, não conhecer o comportamento probabilístico impossibilita o uso do método de máxima verossimilhança na sua forma direta.

O método de máxima verossimilhança utilizado (marginal) consiste em obter os estimadores dos parâmetros por meio da maximização do logaritmo da verossimilhança marginal dos dados, nomeadamente

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.10)$$

Maximizando a verossimilhança dada em (2.10), sendo $\mathbf{V}(\boldsymbol{\theta})$ conhecida, obtém-se os estimadores dos parâmetros de interesse que são dados por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{y} \quad \text{e} \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N}, \quad (3.11)$$

em que $N = \sum_{i=1}^K n_i$. Em geral, $\mathbf{V}(\boldsymbol{\theta})$ é desconhecida e deve ser estimada. Para tais fins, podemos maximizar (2.10) com respeito a $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ simultaneamente. O estimador do vetor $\boldsymbol{\theta}$ tem uma forma fechada para dados completos e balanceados e uma estrutura de covariância de independência condicional para o erro condicional (Graybill, 1976; Laird et al., 1987; Singer e Andrade, 2000). Como em geral, é comum o uso de dados omissos ou

não balanceados, métodos iterativos são necessários para maximizar o logaritmo da verossimilhança restrita marginal. Dentre os métodos de otimização vale ressaltar o método de Newton-Raphson, o método *Scoring* de Fisher e o método EM (Jennrich e Schluchter, 1986; Lindstrom e Bates, 1988).

O problema de maximização não-linear restrita complica muito os algoritmos numéricos. Este problema tem sido abordado por muitos autores e dentre os métodos propostos podemos citar o uso do método da máxima verossimilhança perfilada, ver Demidenko (2004).

Segundo Singer et al. (2012), a metodologia de máxima verossimilhança gera estimadores não-viesados para os efeitos fixos, mas produz estimadores viesados para os parâmetros da matriz de covariâncias intra unidades experimentais por não levar em consideração a perda de graus de liberdade na estimação dos primeiros. Para reduzir o viés desses estimadores, que pode ser grande quando o número de parâmetros de localização é grande em relação ao número de observações, muitos autores, como Laird e Ware (1982), recomendam o uso do método de máxima verossimilhança restrita (MVR), tal método, também conhecido como máxima verossimilhança residual, foi proposto por Patterson e Thompson (1971).

Ainda segundo Singer et al. (2012), seja \mathbf{y} o vetor referente à variável de interesse, a aplicação do método consiste em maximizar a verossimilhança de uma transformação linear do tipo $\mathbf{y}^* = \mathbf{A}^\top \mathbf{y}$ em que \mathbf{A} tem dimensão $[N \times (N - p)]$ e $\mathbb{E}(\mathbf{y}^*) = \mathbf{0}$, ou seja $\mathbf{A}^\top \mathbf{X} = \mathbf{0}$. Assim, para os dados transformados temos que

$$\mathbb{E}(\mathbf{y}^*) = \mathbf{0} \quad \text{e} \quad \mathbb{V}(\mathbf{y}^*) = \mathbf{A}\mathbf{V}(\boldsymbol{\theta})\mathbf{A}^\top.$$

Logo, $\mathbf{y}^* \sim \mathcal{N}_{(N-p)}(\mathbf{0}, \mathbf{A}\mathbf{V}(\boldsymbol{\theta})\mathbf{A}^\top)$. Com base nessas definições, temos que o logaritmo da função de máxima verossimilhança restrita é dado por:

$$l_R(\boldsymbol{\theta}) = -\frac{1}{2} \left\{ \log |\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X}| + \log |\mathbf{V}(\boldsymbol{\theta})| + \hat{\mathbf{e}}^\top \hat{\mathbf{e}} \right\}$$

com $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ e $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{y}$ e maximizando a verossimilhança com relação aos parâmetros, obtém-se estimadores dados por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{y} \quad \text{e} \quad \hat{\sigma}_{REML}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N - p}. \quad (3.12)$$

Contudo, nas situações práticas, as componentes de variância ($\boldsymbol{\theta}$) são desconhecidas. Nesses casos, uma estratégia interessante e conveniente consiste em obter estimativas das

componentes de variância. Dessa forma, deve-se substituir em (2.12) $\mathbf{V}(\boldsymbol{\theta})$ por $\mathbf{V}(\hat{\boldsymbol{\theta}})$. Mais detalhes podem ser obtidos em Demidenko (2004).

3.4.2 BLUE e BLUP

Por meio do teorema de Gauss-Markov para efeitos aleatórios (Harville, 1976) podem ser obtidos o melhor estimador linear não-viesado (*Best Linear Unbiased Estimator*-BLUE) para $\boldsymbol{\beta}$ e o melhor preditor não-viesado (*Best Linear Unbiased Predictor*-BLUP) para o vetor de efeitos aleatórios \mathbf{b} . Diferentes formas de obtenção do BLUP e BLUE, tanto sob o ponto de vista clássico como bayesiano e aplicações podem ser encontradas em Robinson (1991), Searle et al. (1992), Doganaksoy e Balakrishnan (1997), Jiang (1997) ou McCulloch e Searle (2001), por exemplo.

Segundo Rocha (2004), os preditores para os efeitos aleatórios, $\hat{\mathbf{b}}_i$, são importantes, pois, a partir deles, podemos obter os resíduos condicionais, $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{X}}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}$, que são úteis na escolha da estrutura de covariância intra unidades experimentais. Henderson (1975, 1984) obtém conjuntamente os preditores dos efeitos aleatórios e os estimadores para os efeitos fixos por meio da função de distribuição conjunta dos efeitos aleatórios \mathbf{b} e das observações \mathbf{Y} .

Os estimadores dos efeitos fixos e dos preditores para os efeitos aleatórios são encontrados a partir da solução do sistema de equações (2.13) proposto por Henderson (1975), que é uma extensão do método dos mínimos quadrados ordinários, considerando as matrizes \mathbf{D} e $\boldsymbol{\psi}$ conhecidas.

$$\begin{bmatrix} \mathbf{X}^\top \boldsymbol{\psi}^{-1} \mathbf{X} & \mathbf{X}^\top \boldsymbol{\psi}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \boldsymbol{\psi}^{-1} \mathbf{X} & \mathbf{Z}^\top \boldsymbol{\psi}^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \boldsymbol{\psi}^{-1} \mathbf{Y} \\ \mathbf{Z}^\top \boldsymbol{\psi}^{-1} \mathbf{Y} \end{bmatrix},$$

Se as matrizes \mathbf{D} e $\boldsymbol{\psi}$ são desconhecidas, o que normalmente ocorre na prática, deve-se substituí-las pelos seus respectivos estimadores, $\hat{\mathbf{D}}$ e $\hat{\boldsymbol{\psi}}$, na tentativa de encontrar uma aproximação razoável para o BLUE de $\boldsymbol{\beta}$ e para o BLUP de \mathbf{b} . Desse modo, geramos o que chamamos de melhor estimador linear não viciado empírico (EBLUE) e melhor preditor linear não viciado empírico (EBLUP) (Kackar e Harville (1984); Harville e Jeske (1992); Verbeke e Lesaffre (1996b); Jiang (1998)). Dessa forma, o sistema (2.13) é reescrito da seguinte forma:

$$\begin{bmatrix} \mathbf{X}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{X} & \mathbf{X}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{X} & \mathbf{Z}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{Z} + \hat{\mathbf{D}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{Y} \\ \mathbf{Z}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{Y} \end{bmatrix},$$

cuja solução é dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{X}) \mathbf{X}^\top \hat{\mathbf{V}}^{-1} \mathbf{Y} \quad \text{e} \quad \hat{\mathbf{b}} = \hat{\mathbf{D}} \mathbf{Z}^\top \hat{\mathbf{V}}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$

Lembrando que

$$\hat{\mathbf{V}} = \mathbf{Z} \hat{\boldsymbol{\psi}}^{-1} \mathbf{Z}^\top + \hat{\mathbf{D}},$$

implicando

$$\hat{\mathbf{V}}^{-1} = \hat{\boldsymbol{\psi}}^{-1} - \hat{\boldsymbol{\psi}}^{-1} \mathbf{Z} (\mathbf{Z}^\top \hat{\boldsymbol{\psi}}^{-1} \mathbf{Z} + \hat{\mathbf{D}}^{-1})^{-1} + \mathbf{Z}^\top \hat{\boldsymbol{\psi}}^{-1}.$$

No contexto da inferência bayesiana, todos os parâmetros do modelo, sejam fixos (incluindo os componentes de variância) ou aleatórios, são considerados como variáveis aleatórias, conforme o conceito subjetivo de probabilidade (O'Hagan, 1994), pelo qual essa quantifica a incerteza existente sobre eles. Utilizando uma generalização do teorema de Bayes, informações a priori sobre os parâmetros são utilizadas em associação com os dados amostrais (representados no teorema pela função de verossimilhança), possibilitando uma inferência a posteriori sobre aqueles (Box e Tiao, 1973). Se existe informação a priori, a inferência bayesiana pode determinar intervalos de credibilidade mais estreitos que os intervalos de confiança, além de testes mais poderosos. No entanto, é possível trabalhar com as chamadas distribuições a priori não informativas, quando for o caso. A grande vantagem da inferência bayesiana, no contexto dos modelos mistos, é a relativa facilidade de obtenção de desvios-padrão para os elementos de \mathbf{b} (Gianola e Foulley, 1990). Esses autores desenvolveram o método VEIL (Verossimilhança Integrada), que consiste na maximização da densidade conjunta a posteriori dos efeitos, dado \mathbf{y} . As distribuições marginais dos parâmetros, sejam componentes de variância, $\boldsymbol{\beta}$ ou \mathbf{b} , quando maximizadas, fornecem estimadores e preditores pontuais. Quanto aos componentes de variância, esse método é superior ao REML, pois considera a perda de graus de liberdade não só para a estimação de $\boldsymbol{\beta}$ como também dos componentes de variância. Ou seja, a incerteza associada a tais componentes também é levada em conta.

3.4.3 Teste de hipóteses e Critérios de informação

Como pode ser visto em Nobre (2004) e Singer et al. (2012), os testes de interesse são baseados no modelo marginal $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{D}\mathbf{Z}^\top + \boldsymbol{\psi})$ utilizam estatísticas de Wald ou da Razão de Verossimilhanças (RV).

Assintoticamente, sob a hipótese nula, a estatística de Wald tem uma distribuição \mathcal{X}_r^2 com r representando a correspondente redução no número de parâmetros; tal estatística não é apropriada para casos em que o tamanho da amostra é pequeno, a distribuição dos efeitos aleatórios é assimétrica ou a hipótese a ser testada encontra-se na fronteira do espaço paramétrico. Quando o interesse é testar contrastes do tipo $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ com \mathbf{C} representando uma matriz de dimensão $k_1 \times p$, a estatística do teste é

$$\xi_W = (\mathbf{C}\hat{\boldsymbol{\beta}})^\top \left[\mathbf{C}\mathbf{V}(\hat{\boldsymbol{\beta}})\mathbf{C}^\top \right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}}), \quad (3.13)$$

e sua distribuição aproximada é $\mathcal{X}_{\text{posto}(\mathbf{C})}^2$. Dividindo-se (3.13) por $\text{posto}(\mathbf{C})$, obtém-se uma estatística com distribuição aproximada $F_{(\text{posto}(\mathbf{C}), k)}$, com o número de graus de liberdade do denominador k sendo obtido através de aproximação. Diferentes aproximações para k são discutidas em Fai e Cornelius (1996) e em Verbeke e Molenberghs (1997), por exemplo.

O teste da RV pode ser utilizado para testar a hipótese nula de que o modelo com mais parâmetros não se ajusta significativamente melhor do que um modelo restrito (com um número reduzido de parâmetros). A estatística da RV é dada por

$$\xi_{RV} = -2(L_1 - L_2) \quad (3.14)$$

com L_1 representando o máximo da log-verossimilhança sob o modelo restrito (encaixado) e L_2 a respectiva função correspondente do modelo com r parâmetros adicionais. Quando o modelo reduzido não se situa na fronteira do espaço paramétrico, tem-se que $\xi_{RV} \sim \mathcal{X}_r^2$. Self e Liang (1987) mostram que quando o modelo reduzido se situa na fronteira do espaço paramétrico, então a distribuição assintótica de (3.14) é uma mistura de distribuições \mathcal{X}^2 . O teste da RV não é apropriado para testar hipóteses referentes aos efeitos fixos quando se utiliza a log-verossimilhança restrita, uma vez que ela exclui tais efeitos. Recentemente Verbeke e Molenberghs (2003) utilizaram o teste “Score” e observaram os mesmos “problemas” dos testes de Wald e da RV.

Quando os modelos não são encaixados ou quando a hipótese de interesse situa-se na fronteira do espaço paramétrico, podem-se utilizar alguns critérios de informação fundamentados na teoria da decisão que penalizam os modelos com um grande número de parâmetros. Alguns desses critérios são baseados nas estatísticas AIC (Akaike Information Criterion), o BIC (Bayesian Information Criterion) e o CAIC (Consistent Akaike’s

Information Criterion) definidos como

$$\text{AIC} = -2l + 2d$$

$$\text{BIC} = -2l + n \ln n$$

$$\text{CAIC} = -2l + d(\ln n + 1)$$

com l representando o máximo da log-verossimilhança (completa ou restrita), d o número de parâmetros do modelo e n o número de observações. Quanto menor for o valor dessas estatísticas, maior evidência favorável ao modelo em questão. Detalhes sobre testes de hipóteses e critérios de seleção para modelos lineares mistos podem ser encontrados em Bozdogan (1987), Andreoni (1989), Öfversten (1993), Stram e Lee (1994), Suyama (1995), Christensen (1996), Verbeke e Molenberghs (1997), Pinheiro e Bates (2000), dentre outros.

4 ANÁLISE DE DIAGNÓSTICO

4.1 Modelo normal linear

Segundo Lobato Júnior (2005), modelos estatísticos são geralmente descrições aproximadas de processos bastante complexos, conseqüentemente podem levar a resultados imprecisos, daí surge uma importante motivação para o estudo de técnicas que avaliem essa imprecisão. Um dos objetivos da avaliação de modelos ajustados, visa a detecção de observações que apresentem um comportamento diferenciado das demais. Tais observações podem ser classificadas segundo Paula (2010), do seguinte modo:

- **Pontos Aberrantes:** são pontos que tem uma certa influência sobre os valores ajustados, embora não estejam muito afastados dos demais pontos.
- **Pontos de Alavanca:** são pontos que estão mais afastados dos pontos no subespaço gerado pelas colunas da matriz \mathbf{X} . Estes pontos não influenciam de forma significativa as estimativas dos parâmetros, porém fazem com que as variâncias dos valores ajustados dos pontos próximos a ele sejam maiores do que as variâncias dos valores ajustados correspondentes aos demais pontos.
- **Pontos Influentes:** são combinações dos pontos aberrantes e de alavanca.

4.1.1 Análise de Resíduos

Segundo Singer et al. (2012), os resíduos podem ser utilizados, no caso de modelos de regressão normal linear, para verificar homoscedasticidade, existência de pontos discrepantes, normalidade e independência dos erros, como veremos a seguir.

O resíduo para a i -ésima observação é definido como uma função do tipo $r_i = r(y_i; \hat{\mu}_i)$, cujo o objetivo é medir a discrepância entre o valor observado e o valor ajustado da i -ésima

observação. A definição mais usual é a de resíduo ordinário dada por:

$$r_i = y_i - \hat{\mu}_i.$$

Para $\mathbf{r} = (r_1, r_2, \dots, r_n)^\top$, temos que, para o modelo de regressão normal linear,

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y}, \end{aligned}$$

em que \mathbf{H} é a matriz *hat*, $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. O i -ésimo elemento da diagonal da matriz \mathbf{H} , denotado por h_{ii} , desempenha um importante papel na construção de técnicas de diagnóstico. Sabendo que $\text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii}$ e partindo do pressuposto que todos os pontos exerçam a mesma influência sobre os valores ajustados, espera-se que o valor de h_{ii} esteja próximo de p/n . Convém então examinar aqueles pontos tais que $h_{ii} \geq 2p/n$, que podem ser diagnosticados como pontos de **alavanca** e geralmente estão localizados em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} , Paula (2010). Em relação ao modelo de regressão normal linear, temos que $r_i \sim \mathcal{N}(0, (1 - h_{ii})\sigma^2)$. Assim pode-se obter os **resíduos studentizados**, denotados por t_i , dividindo-se cada r_i pelo seu respectivo desvio padrão amostral $s(1 - h_{ii})^{1/2}$, em que $s^2 = \sum_{i=1}^n \frac{r_i^2}{(n - p)}$, é o estimador da variância σ^2 . Logo,

$$t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}, i = 1, \dots, n.$$

Como r_i não é independente de s^2 , t_i não segue uma distribuição *t* de *Student*, como se poderia esperar. Paula (2010) mostra que o problema da dependência entre r_i e s^2 pode ser contornado, substituindo s^2 por $s^2_{(i)}$ na expressão de t_i anterior, onde $s^2_{(i)}$ é o erro quadrático médio correspondente ao modelo sem a i -ésima observação, dado por:

$$s^2_{(i)} = s^2 \left(\frac{n - p - 1}{n - p - t^2_{(i)}} \right).$$

Dessa forma, temos que o novo resíduo *studentizado* é dado pela expressão,

$$t_i^* = t_i \left(\frac{n - p - 1}{n - p - t^2_{(i)}} \right)^{1/2}.$$

4.1.2 Análise de Sensibilidade

Ainda segundo Singer et al. (2012), a análise de sensibilidade visa avaliar o comportamento do ajuste de um modelo quando ele está sujeito a algum tipo de perturbação, ou seja, sob alguma mudança nas hipóteses ou nos dados. Como cada observação não tem a mesma influência em todas as características do ajuste do modelo, é natural que se defina aquela na qual se quer focar a análise. Se o objetivo for fazer previsões, então é razoável medir a influência das observações nos valores preditos e não nos parâmetros de localização, como mencionam Chatterjee e Hadi (1986, 1988). Dentre as abordagens mais utilizadas na prática para medir influência em modelos lineares, destacam-se aquelas baseadas em influência local considerada em Cook (1986) e aquelas obtidas por intermédio da eliminação de observações (influência global).

4.1.2.1 Influência

Considere a função de verossimilhança para o parâmetro β a seguir:

$$l_{\mathbf{w}}(\beta, \mathbf{y}) = \sum_{i=1}^n w_i l(\beta, \mathbf{y}_i)$$

em que $l(\beta, \mathbf{y}_i)$ denota o logaritmo da função de verossimilhança correspondente à i -ésima observação e w_i é um tipo de perturbação, tal que $0 \leq w_i \leq 1$. Quando $w_1 = w_2 = \dots = w_n = 1$, significa que não há perturbação no modelo e quando $w_i = 0$ significa que a i -ésima observação foi excluída.

A medida de influência mais conhecida é a distância de Cook, definida por:

$$D_{\mathbf{w}} = \frac{(\hat{\beta} - \hat{\beta}_{\mathbf{w}})^{\top} (\mathbf{X}^{\top} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{\mathbf{w}})}{ps^2}$$

em que $\hat{\beta}_{\mathbf{w}} = (\mathbf{X}^{\top} \Delta \mathbf{X})^{-1} \mathbf{X}^{\top} \Delta \mathbf{y}$ e $\Delta = \text{diag}\{w_1, w_2, \dots, w_n\}$. Esta medida mede quanto a perturbação $\mathbf{w} = (w_1, w_2, \dots, w_n)^{\top}$ afasta $\hat{\beta}_{\mathbf{w}}$ de $\hat{\beta}$ segundo a métrica $\mathbf{X}^{\top} \mathbf{X}$. Quando o i -ésimo ponto é excluído, Paula (2010) mostra que a distância de Cook fica expressa por:

$$\begin{aligned} D_i &= \frac{(\hat{\beta} - \hat{\beta}_{(i)})^{\top} (\mathbf{X}^{\top} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} \\ &= t_i^2 \frac{h_{ii}}{(1 - h_{ii})p} \end{aligned}$$

Portanto, D_i será grande quando o i -ésimo ponto for influente (t_i grande) e/ou quando

h_{ii} for próximo de um.

4.1.2.2 Influência Local

O método de influência local é considerado um dos mais modernos dentro da análise de diagnóstico (Paula, 2010). Proposto por Cook (1986) consiste em verificar a existência de pontos que, sob modificações infinitesimais no modelo, causam variações desproporcionais nos resultados do ajuste. Considerando o caso geral da distância de Cook $D_{\mathbf{w}}$ definida anteriormente, queremos estudar as mudanças produzidas nesta medida quando $\mathbf{w}_i \rightarrow 1, \forall i$. Expandindo $D_{\mathbf{w}}$ em série de Taylor até segunda ordem, em torno de $\mathbf{w}_0 = 1$, temos que:

$$D_{\mathbf{w}} \cong D_{\mathbf{w}_0} + (\mathbf{w}_0 - \mathbf{w})^\top D'_{\mathbf{w}_0} + \frac{1}{2}(\mathbf{w}_0 - \mathbf{w})^\top D''_{\mathbf{w}_0}(\mathbf{w}_0 - \mathbf{w}).$$

Pode-se mostrar que

$$D''_{\mathbf{w}_0} = \text{diag}(\mathbf{r})\mathbf{H}\text{diag}(\mathbf{r}),$$

em que $\text{diag}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$. Cook estuda a maior variação de $D_{\mathbf{w}}$ em torno de \mathbf{w}_0 , o que equivale a maximizar a forma quadrática

$$\mathbf{l}^\top \mathbf{F} \mathbf{l},$$

em que $\mathbf{l} = \mathbf{w}_0 - \mathbf{w}$, $\mathbf{l}^\top \mathbf{l} = 1$ e $\mathbf{F} = \text{diag}(\mathbf{r})\mathbf{H}\text{diag}(\mathbf{r})$.

O máximo da forma quadrática $\mathbf{l}^\top \mathbf{F} \mathbf{l}$ corresponde ao maior autovalor da matriz \mathbf{F} , denotado por λ_{max} . As coordenadas do autovetor \mathbf{l}_{max} , associado a λ_{max} , contém as influências locais das observações na direção que cause maior alteração. Assim, o gráfico de $|\mathbf{l}_{max}|$ contra a ordem das observações pode revelar os pontos com maior influência na vizinhança de \mathbf{w}_0 , estes pontos podem causar mudanças substanciais nas estimativas dos parâmetros sob pequenas perturbações no modelo.

4.2 Modelos Lineares Mistos

4.2.1 Análise de Resíduos

Segundo Nobre e Singer (2007), como os modelos lineares mistos tem duas fontes de variação (\mathbf{e} e \mathbf{b}), tipos diferentes de resíduos podem ser definidos e a análise correspondente é mais complexa. Em modelos lineares mistos existem três tipos de resíduos:

1. **Resíduos marginais:** $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, que é o preditor dos erros marginais $\boldsymbol{\xi} = \mathbf{y} - \mathbb{E}(\mathbf{y}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.
2. **Resíduos condicionais:** $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}$ que é o preditor dos erros condicionais $\mathbf{e} = \mathbf{y} - \mathbb{E}(\mathbf{y}|\mathbf{b}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$.
3. O BLUP, $\mathbf{Z}\hat{\mathbf{b}}$, que é o preditor dos efeitos aleatórios, $\mathbf{Z}\mathbf{b} = \mathbb{E}(\mathbf{y}|\mathbf{b}) - \mathbb{E}(\mathbf{y})$.

Segundo Pinheiro e Bates (2000), antes de quaisquer inferências, duas suposições devem ser verificadas nos modelos lineares mistos: se os erros intra unidades experimentais são independentes e identicamente distribuídos seguindo uma distribuição normal com média zero e variância σ^2 e se são independentes dos efeitos aleatórios. A outra suposição refere-se à normalidade dos efeitos aleatórios e são independentes para as diferentes unidades experimentais.

No intuito de verificar o afastamento dessas suposições, Pinheiro e Bates (2000) propõem o uso do gráfico de probabilidade normal dos resíduos condicionais para avaliar a suposição de normalidade e o gráfico dos resíduos condicionais *versus* os valores ajustados para avaliar a suposição de homoscedasticidade. Além disso, os resíduos condicionais também podem ser utilizados para identificação de pontos discrepantes. Entretanto, Nobre (2004), com base na possibilidade dos elementos de $\hat{\mathbf{e}}$ apresentarem variâncias diferentes, propõe uma padronização dos resíduos condicionais dada por:

$$\hat{\mathbf{e}}_i^* = \frac{\hat{\mathbf{e}}_i}{\sigma\sqrt{q_{ii}}},$$

em que $\hat{\mathbf{e}}_i$ é o i -ésimo elemento de $\hat{\mathbf{e}}$ e q_{ii} é o i -ésimo elemento da matriz \mathbf{Q} . A matriz \mathbf{Q} é definida como, $\mathbf{Q} = \mathbf{M} - \mathbf{M}\mathbf{X}(\mathbf{X}^\top\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M}$, em que $\mathbf{M} = \sigma^2\mathbf{V}^{-1} = (\mathbf{Z}\mathbf{D}\mathbf{Z}^\top + \boldsymbol{\psi})^{-1}$.

Segundo Costa (2010), o resíduo condicional padronizado, $\hat{\mathbf{e}}^*$, é mais eficaz na verificação de possível afastamento da suposição de homoscedasticidade, fazendo uso do gráfico de resíduos condicionais padronizados *versus* valores ajustados.

Quando o interesse é verificar a suposição de normalidade para o erro \mathbf{e} , não é aconselhável utilizar $\hat{\mathbf{e}}$, porque ele é confundido por \mathbf{b} ; logo, quando \mathbf{b} se afasta muito da normalidade, $\hat{\mathbf{e}}$ pode não apresentar características de normalidade, mesmo quando \mathbf{e} segue uma distribuição normal.

Sendo assim, Hilden-Minton (1995) citado por Nobre (2004) define a fração de confundimento de $\hat{\mathbf{e}}_i$ como:

$$CF_i = 1 - \frac{\mathbf{U}_i^\top \mathbf{Q} \mathbf{U}_i}{\mathbf{U}_i^\top \mathbf{Q} \mathbf{U}_i}$$

em que \mathbf{U}_i é a i -ésima coluna da matriz \mathbf{I}_n . A medida CF_i é interpretada como sendo a proporção de variabilidade de $\hat{\mathbf{e}}_i$ quando confundida por $\hat{\mathbf{b}}$. É um valor entre zero e um e quanto mais próximo de um for o CF_i maior é o confundimento. Dessa forma, o autor propõe utilizar uma transformação linear, $\mathbf{L}^\top \hat{\mathbf{e}}$, de tal forma que $\mathbf{L}^\top \hat{\mathbf{e}}$ tenha confundimento mínimo. Denotando a i -ésima coluna da matriz \mathbf{L} por \mathbf{l}_i , $\mathbf{l}_i^\top \hat{\mathbf{e}}$ é dito ser pouco confundido se:

$$\pi_i = \frac{\mathbf{l}_i^\top \mathbf{I}_n \mathbf{Q} \mathbf{I}_n \mathbf{Q} \mathbf{l}_i}{\mathbf{l}_i^\top \mathbf{I}_n \mathbf{Q} \mathbf{I}_n \mathbf{l}_i}$$

é máximo, sujeito a restrição que $\text{Var}(\mathbf{l}_i^\top \hat{\mathbf{e}}) \propto \mathbf{l}_i^\top \mathbf{I}_n \mathbf{Q} \mathbf{I}_n \mathbf{l}_i > 0$. Assim o resíduo com confundimento mínimo, é obtido pela maximização da expressão acima gerando assim, uma sequência de erros não correlacionados homoscedásticos com variância σ^2 e fração de confundimento mínimo, seu uso é mais eficiente na verificação da normalidade dos erros condicionais por intermédio do gráfico de probabilidade normal com envelope simulado.

Cada tipo de resíduo é útil para avaliar algumas suposições do modelo (2.3), assim como sintetizado na Tabela 4, adaptada por Nobre e Singer (2007).

Tabela 4: Uso dos diferentes resíduos na análise de diagnóstico.

Diagnóstico para	Tipo de resíduo	Gráfico
Linearidade dos efeitos $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$	Marginal	$\hat{\boldsymbol{\xi}}_i$ vs. Variáveis explicativas
Matriz de Covariância (\mathbf{V}_i)	Marginal	\mathbf{v} vs. Índices dos indivíduos
Presença de observações discrepantes	Condicional	$\hat{\mathbf{e}}_i^*$ vs. Índices das observações
Homoscedasticidade dos erros condicionais (\mathbf{e}_i)	Condicional	$\hat{\mathbf{e}}_i^*$ vs. Valores ajustados
Normalidade dos erros condicionais (\mathbf{e}_i)	Condicional	QQ plot para os resíduos de confundimento mínimo ou resíduos do modelo linear com bandas de tolerância
Presença de indivíduos discrepantes	EBLUP	\mathbf{M}_i vs. Índices dos indivíduos
Normalidade dos efeitos aleatórios (\mathbf{b}_i)	EBLUP	QQ plot ponderado $\hat{\mathbf{b}}_i$

Fonte: Nobre e Singer (2007)

4.2.2 Análise de Sensibilidade

4.2.2.1 Eliminação de observações

Cook (1977) apresentou o método de eliminação de observações com o intuito de identificar possíveis pontos que estivessem influenciando, de forma desproporcional, na estimatimação dos parâmetros. Nesta direção, há o que é denominado de fórmulas de atualização que visam atualizar as estimativas dos parâmetros quando uma determinada observação for excluída. Nos modelos lineares mistos a eliminação de observações não se apresenta como uma tarefa simples dado que a estimação é realizada iterativamente, sendo necessário aproximações para as fórmulas de atualização. A preocupação acerca da eliminação de observações, segundo Cook e Weisberg (1980), é fundamentada na possibilidade de que, ao se excluir uma determinada observação e analisar sua influência nas estimativas, a influência causada por um conjunto de observações seja mascarada e com isso observações, que conjuntamente com outras estejam influenciando as estimativas, não sejam identificadas.

No que se trata da análise da influência de um conjunto de observações tomando a ideia da eliminação desse conjunto, como pode ser visto em vários textos em muitas estruturas de modelos distintas, a distância de Cook é a medida mais utilizada nesta tarefa.

A distância de Cook para modelos lineares misto é dada por:

$$D_I = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I)^\top (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_I)}{c} = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_I)^\top \mathbf{V}^{-1}(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_I)}{c}, \quad (4.1)$$

em que c é um parâmetro de escala e $\hat{\boldsymbol{\beta}}_I$ e $\hat{\mathbf{Y}}_I$ são as estimativas de $\boldsymbol{\beta}$ e \mathbf{Y} retirando o conjunto de observações I .

Segundo Fung et al. (2002), a equação (3.1) pode ser utilizada para detectar tanto observações como indivíduos influentes, pois uma vez que I representa um conjunto de observações e fazendo I todas as observações de um determinado indivíduo, tem-se exatamente a influência causada por este. Entretanto, Banerjee (1998) e Tan et al. (2001) discutem limitações da distância de Cook dada na forma (3.1) na classe de modelos lineares mistos, pois tal medida pode não detectar observações ou indivíduos influentes quando estes apresentam grande influência em $\hat{\mathbf{b}}$.

No sentido de evitar a não detecção de possíveis conjuntos de observações influentes, Tan et al. (2001) propuseram a Distância de Cook Condicional que tem como fundamento

o uso do BLUE e do BLUP, sendo esta medida dada por:

$$D_i^{cond} = \sum_{j=1}^K \frac{\mathbf{P}_{j(i)}^\top \mathbf{P}_{j(i)}}{w},$$

em que $\mathbf{P}_{j(i)} = (\mathbf{X}_j \hat{\boldsymbol{\beta}} + \mathbf{Z}_j \hat{\mathbf{b}}_j) - (\mathbf{X}_j \hat{\boldsymbol{\beta}}_{(i)} + \mathbf{Z}_j \hat{\mathbf{b}}_{j(i)})$, sendo $\hat{\boldsymbol{\beta}}_{(i)}$ e $\hat{\mathbf{b}}_{j(i)}$ representam, respectivamente, $\hat{\boldsymbol{\beta}}$ e \mathbf{b}_j obtidos quando eliminamos da amostra a i -ésima observação e $w = \sigma^2[(N-1)c + p]$.

A distância de Cook condicional definida é útil para verificação de observações que estejam influenciando as estimativas dos efeitos fixos e aleatórios. Nesse contexto, a equação acima pode ser decomposta em três partes permitindo abordar a influência em partes distintas do modelo ajustado. A decomposição é dada pela seguinte expressão:

$$D_i^{cond} = D_{1i}^{cond} + D_{2i}^{cond} + D_{3i}^{cond},$$

em que

$$D_{1i}^{cond} = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i)^\top (\mathbf{X}^\top \mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i)}{w} = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_i)^\top (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_i)}{w},$$

$$D_{2i}^{cond} = \frac{(\hat{\mathbf{b}} - \hat{\mathbf{b}}_i)^\top \mathbf{Z}^\top \mathbf{Z}(\hat{\mathbf{b}} - \hat{\mathbf{b}}_i)}{w},$$

$$D_{3i}^{cond} = \frac{2(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i)^\top \mathbf{X}^\top \mathbf{Z}(\hat{\mathbf{b}} - \hat{\mathbf{b}}_i)}{w},$$

Segundo Nobre (2004), o primeiro termo, D_{1i}^{cond} , é uma medida útil para avaliar a influência da i -ésima observação em $\hat{\boldsymbol{\beta}}$, não incluindo a padronização pela matriz de covariâncias de \mathbf{Y} . O segundo termo, D_{2i}^{cond} , é uma medida útil para avaliar a influência da i -ésima observação em $\hat{\mathbf{b}}$. Já o terceiro termo, D_{3i}^{cond} , é uma medida de covariância, entre a mudança nas estimativas do BLUE e BLUP, quando eliminamos a i -ésima observação.

Contudo, ao excluir o conjunto de observações referentes a um indivíduo, tem-se que a obtenção de alguns BLUP não é possível. Com base nesse problema, Nobre (2004) propõe utilizar a médias das distâncias ao invés da distância de Cook condicional definida

que é dada por:

$$D_{i.}^{cond} = (n_i)^{-1} \sum_{j \in I} D_j^{cond},$$

com com I representando o conjunto das n_i observações da i -ésima unidade experimental.

4.2.2.2 Pontos de Alavanca

No contexto dos modelos mistos, assumindo \mathbf{V} conhecida, vamos definir a matriz de alavancagem generalizada para os valores marginais ajustados, assim como definiram Wei et al. (1998), da seguinte forma:

$$\begin{aligned} \mathbf{L}_1 &= \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} \\ &= \frac{\partial \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}}{\partial \mathbf{y}^\top} \\ &= \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}, \end{aligned}$$

correspondendo assim, à matriz de projeção oblíqua do subespaço formado pelas colunas de \mathbf{X} usando a distância induzida por \mathbf{V}^{-1} e fornecendo informações sobre o grau de alavancagem das observações com respeito aos valores marginais ajustados, como pode ser observado em Fung et al. (2002) e Demidenko (2004). Por esta razão denomina-se de *matriz de alavancagem generalizada marginal*.

A i -ésima unidade amostral da matriz de alavancagem generalizada marginal é a i -ésima matriz bloco diagonal de \mathbf{L}_1 , ou seja

$$\mathbf{L}_{1i} = \mathbf{X}_i(\mathbf{X}_i^\top \mathbf{V}^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^\top \mathbf{V}_i^{-1}.$$

Uma vez que (i) possamos estar interessados em detectar o alto de grau de alavancagem de observações pertencentes a uma mesma unidade amostral e (ii) há casos em que o alto grau de alavancagem de uma certa observação não implica no alto grau de alavancagem da unidade amostral correspondente, vamos definir a j -ésima observação da i -ésima unidade amostral da matriz de alavancagem generalizada marginal como o j -ésimo elemento da diagonal principal de \mathbf{L}_{1i} , denotando-o como $\mathbf{L}_{1i(jj)}$.

Assim como definido em Nobre e Singer (2011), altos valores de $\mathbf{L}_{1i(jj)}$ indicam alto poder de alavancagem da j -ésima observação pertencente a i -ésima unidade amostral com relação aos valores marginais ajustados. Embora, na prática, observações ou unidades experimentais com alta alavancagem são definidas de maneira subjetiva (por inspeção

visual dos gráficos de alavanca), alguns critérios objetivos tem sido sugeridos na literatura. Por exemplo, dado que:

$$tr(\mathbf{L}_1) = \sum_{i=1}^K \sum_{j=1}^{n_i} \mathbf{L}_{1i(jj)} = p,$$

pode-se considerar a j -ésima observação da i -ésima unidade amostral como sendo de alta alavancagem para os valores marginais ajustados se $\mathbf{L}_{1i(jj)} \geq \frac{2p}{N}$ e considerar a i -ésima unidade amostral como sendo de alta alavancagem para os valores marginais ajustados se

$$\frac{tr(\mathbf{L}_1)}{n_i} \geq \frac{2p}{N},$$

como sugerido em Demidenko (2004).

Segundo Nobre e Singer (2011), uma observação pode afetar os valores marginais e condicionais ajustados, é conveniente avaliar a influência conjunta de cada observação em ambos. Vamos usar $\hat{\mathbf{y}}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{b}}$, ao invés de $\hat{\mathbf{y}}$ e vamos definir a matriz de alavancagem generalizada conjunta como

$$\begin{aligned} \mathbf{L} &= \frac{\partial \hat{\mathbf{y}}^*}{\partial \mathbf{y}^\top} = \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}^\top} + \frac{\partial \mathbf{Z}\hat{\mathbf{b}}}{\partial \mathbf{y}^\top} \\ &= \mathbf{L}_1 + \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q} \\ &= \mathbf{L}_2 \mathbf{V}^{-1} + [\mathbf{I}_N - \mathbf{L}_2 \mathbf{V}^{-1}] \mathbf{L}_1, \end{aligned}$$

em que \mathbf{L}_1 é a matriz de alavancagem generalizada marginal, $\boldsymbol{\Gamma} = \mathbf{I}_K \otimes \mathbf{D}$ e

$$\mathbf{L}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top.$$

A matriz \mathbf{L}_2 representa o quanto da variabilidade intra unidades experimentais está sendo explicada pela presença dos efeitos aleatórios.

Nobre (2004) e Demidenko e Stukel (2005) sugerem que

$$\mathbf{H}_2 = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top \mathbf{Q},$$

poderia ser usada como uma matriz de alavancagem generalizada para o componente de efeitos aleatórios dos valores ajustados condicionais. Dado que

$$\mathbf{H}_2 = \mathbf{L}_2 \mathbf{V}^{-1} + [\mathbf{I}_N - \mathbf{L}_1],$$

depende da matriz de alavancagem generalizada marginal, os efeitos de ambos os componentes de efeitos fixos e aleatórios dos valores ajustados condicionais podem ser confun-

dados.

A melhor alternativa para mensurar o grau de alavancagem de observações e de indivíduos nos componentes de efeitos aleatórios dos valores ajustados condicionais é usar a *matriz de alavancagem generalizada do componente aleatório* \mathbf{L}_2 ao invés de \mathbf{H}_2 . Observações que apresentam alto poder de alavancagem com relação ao componente de efeitos aleatórios dos valores ajustados condicionais deve exercer um peso desproporcional na estimativa dos componentes da variância do modelo. Assim como está definido em Nobre e Singer (2011), vamos definir a i -ésima unidade amostral da matriz de alavancagem generalizada do componente aleatório como $\mathbf{L}_{2i} = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$. A matriz de alavancagem generalizada do componente aleatório correspondente a j -ésima observação da i -ésima unidade amostral é $\mathbf{L}_{2i(jj)}$ que por sua vez corresponde ao j -ésimo elemento da diagonal principal de \mathbf{L}_{2i} . A i -ésima unidade amostral é dita ter alto poder de alavancagem com relação ao componente aleatório dos valores ajustados condicionais se $(n_i)^{-1} \text{tr}(\mathbf{L}_{2i})$ é “grande” (por exemplo, se $(n_i)^{-1} \text{tr}(\mathbf{L}_{2i}) \geq 2 \text{tr}(\mathbf{L}_2)/N$) comparado com outros valores; similarmente, A j -ésima observação pertencente a i -ésima unidade amostral é dita ter alto poder de alavancagem com relação ao componente aleatório dos valores ajustados condicionais se $\mathbf{L}_{2i(jj)}$ é “grande” (por exemplo, $\mathbf{L}_{2i(jj)} \geq 2 \text{tr}(\mathbf{L}_2)/N$).

No capítulo 5, serão aplicadas as técnicas aqui apresentadas.

5 *ABORDAGEM BAYESIANA*

5.1 Introdução

A estatística clássica é fundamentada no crêdulo de que toda a informação a respeito do fenômeno aleatório que se deseja compreender é oriunda dos dados a serem obtidos pelo pesquisador, ou seja, da amostra. Sendo assim, toda e qualquer inferência é embasada apenas naquilo que se observou, seja através da obtenção de uma amostra ou da realização de um experimento, ou naquilo que poderia ser observado caso o experimento fosse repetido ou uma nova amostra fosse observada. Em um exemplo prático, de acordo com o pensamento clássico, pode-se imaginar que o efeito médio que a aplicação de certo tratamento em pacientes gravemente doentes tem sobre o tempo de sobrevivência dos mesmos é uma quantidade específica, fixa, porém desconhecida.

A filosofia bayesiana, em contrapartida, acredita que os parâmetros de interesse, por serem desconhecidos, se tratam de quantidades aleatórias. Desta forma, a especificação de distribuições de probabilidade para os parâmetros é natural. Em exemplo análogo ao anteriormente apresentado, o pensamento bayesiano considera que o efeito médio é dotado de variabilidade, propondo que, a incerteza que se tem a seu respeito deva ser representada por uma distribuição de probabilidade. É interessante proceder a associação deste pensamento à dados que envolvem opinião pública, doenças, entre outros.

Por outro ponto de vista, então, os bayesianos consideram que não só os dados obtidos são importantes para o processo inferencial, como também a informação, com correspondente grau de incerteza (precisão), que o pesquisador carrega a respeito daquilo que está estudando. Sendo assim, é aberto espaço para que o conhecimento e a experiência do pesquisador no assunto possam contribuir com as inferências a serem realizadas. Este conhecimento do pesquisador sobre os parâmetros de interesse é dito “conhecimento à priori” e será representado pela “distribuição à priori”.

A informação à priori (distribuição à priori) será combinada com a informação presente nos dados obtidos (função de verossimilhança) de modo que aquele que carregar um maior

grau de certeza seja mais valorizado. O resultado final desta combinação de fontes de informação é representado pela distribuição à posteriori e será a base para toda e qualquer inferência que se fizer interessante.

5.2 Teorema de Bayes

O teorema de Bayes é uma proposição sobre probabilidades condicionais indiscutível desde que se aceitem os axiomas de Kolmogorov (Box e Tiao, 1973) para definir medidas de probabilidade. O que tem dado lugar a grande controvérsia é a sua interpretação e sua aplicação a problemas de inferência estatística. Ocupando, como é óbvio, lugar crucial na inferência bayesiana.

Seja $\boldsymbol{\theta}$ uma quantidade de interesse desconhecida cujos possíveis valores são pertencentes ao espaço paramétrico Θ . O objetivo da inferência bayesiana pode ser a estimação de $\boldsymbol{\theta}$ ou o teste de alguma hipótese envolvendo valores de $\boldsymbol{\theta}$. Um dos principais ingredientes para a realização de inferência bayesiana é a distribuição a posteriori, que representa o conhecimento a respeito de $\boldsymbol{\theta}$ após a observação dos dados \mathbf{x} .

A distribuição a posteriori é obtida através do teorema de Bayes, isto é, definido da seguinte forma (Box e Tiao, 1973):

Suponha que $\mathbf{x}^\top = (x_1, \dots, x_n)$ é um vetor de n observações cuja distribuição de probabilidade $p(\mathbf{x}|\boldsymbol{\theta})$ depende dos valores de k parâmetros $\boldsymbol{\theta}^\top = (\theta_1, \dots, \theta_k)$. Suponha ainda que $\boldsymbol{\theta}$ em si tem uma distribuição de probabilidade $p(\boldsymbol{\theta})$. Então,

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})},$$

em que $p(\boldsymbol{\theta})$ e $p(\boldsymbol{\theta}|\mathbf{x})$ são, respectivamente, a distribuição a priori e a posteriori de $\boldsymbol{\theta}$.

As informações sobre o vetor de parâmetros de interesse, $\boldsymbol{\theta}$, que são obtidas antes da realização de um experimento podem ser representadas probabilisticamente através das chamadas distribuições a priori. Portanto, essas distribuições expressam o estado de conhecimento ou ignorância sobre o vetor de parâmetros $\boldsymbol{\theta}$ antes dos dados serem obtidos (Box e Tiao, 1973).

A distribuição a posteriori de um parâmetro $\boldsymbol{\theta}$ contém toda a informação probabilística a respeito deste parâmetro e um gráfico da sua função de densidade a posteriori é a melhor descrição do processo de inferência. No entanto, algumas vezes é necessário resumir a informação contida nesta posteriori através de alguns poucos valores numéricos. O caso

mais simples é a predição pontual de θ onde se resume a distribuição a posteriori através de um único elemento, $\hat{\theta}$.

5.3 Informações a priori

A distribuição a priori representa o conhecimento prévio do pesquisador antes de se observar os dados. De acordo com Leandro (2001) devido, em geral, a subjetividade da escolha das densidades a priori, estatísticos frequentistas se opõem ao uso desta informação adicional e se sustentam no fato de que a função densidade de probabilidade a posteriori ser bastante sensível a escolha de prioris distintas. Porém, os estatísticos bayesianos tem apoio em alguns argumentos, por exemplo este feito por Gelman et. al. (1997): “Todos os métodos estatísticos que usam probabilidades são subjetivos no sentido que se baseiam em idealizações matemáticas do mundo”.

5.3.1 Distribuições a priori subjetivas

A distribuição a priori subjetiva é usada quando não existe nenhuma base teórica objetiva para o ajuste do modelo. Nessas situações, a distribuição a priori representa a expectativa subjetiva do pesquisador a respeito do parâmetro de interesse. Com isso, a distribuição a posteriori para o parâmetro, obtida combinando essa distribuição a priori e a função de verossimilhança, mostra como o conhecimento pessoal do pesquisador é modificado pela informação dos dados (Box e Tiao, 1973). O problema de como quantificar essa informação subjetiva e transformá-la em uma distribuição a priori que possa ser usada na abordagem bayesiana, tem sido discutido abundantemente na literatura. Uma interação entre o pesquisador e o estatístico é fundamental para a determinação de tal a priori.

5.3.2 Distribuições a priori conjugadas

A ideia de distribuição conjugada é que as distribuições a priori e a posteriori pertençam a mesma classe de distribuições e assim a atualização do conhecimento que se tem sobre o parâmetro θ envolve apenas uma mudança nos parâmetros indexadores da família de distribuições a priori, denominados hiperparâmetros, que diferem do parâmetro θ . Neste caso, o aspecto sequencial do método bayesiano pode ser explorado definindo-se apenas a regra de atualização dos hiperparâmetros já que as distribuições permanecem

as mesmas. Pode-se definir a distribuição a priori conjugada da seguinte forma (Ehlers, 2007):

- Se $F = \{p(\mathbf{x}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ é uma classe de distribuições amostrais então uma classe de distribuições P é conjugada a F se

$$\forall p(\mathbf{x}|\boldsymbol{\theta}) \in F \quad \text{e} \quad p(\boldsymbol{\theta}) \in P \Rightarrow p(\boldsymbol{\theta}|\mathbf{x}) \in P.$$

Gamerman e Lopes (2006) ressaltam que a distribuição a priori conjugada deve ser usada com cuidado, pois sua utilização está muitas vezes associadas às facilidades analíticas nem sempre é uma representação adequada do conhecimento a priori do parâmetro.

5.3.3 Distribuições a priori não-informativas

A priori não-informativa refere-se ao caso em que pouca ou nenhuma informação é disponível a priori, isto é, a informação a respeito dos parâmetros do modelo não é significativa se comparada à informação proveniente da amostra.

Na literatura bayesiana, há várias razões para se utilizar prioris não-informativas, dentre elas:

1. É possível, de imediato, fazer uma análise bayesiana para o modelo, uma vez que existem inúmeras prioris não-informativas propostas na literatura;
2. As prioris informativas, baseadas em métodos subjetivos, são frequentemente mais difíceis de aplicar na prática;
3. Ao formular um procedimento estatístico alternativo ou para um estudo comparativo entre metodologias estatísticas, este seria mais facilmente aceito se utilizássemos prioris não-informativas;
4. Se há uma quantidade substancial de dados, a distribuição a priori terá pouco efeito. Isto justificaria o uso de prioris não informativas.

5.3.3.1 Priori Uniforme

A primeira ideia de não informação a priori que se pode ter é pensar em todos os possíveis valores de $\boldsymbol{\theta}$ como igualmente prováveis, isto é, com uma distribuição a priori uniforme. Neste caso, fazendo

$$p(\boldsymbol{\theta}) \propto k$$

para $\boldsymbol{\theta}$ variando em um subconjunto da reta, significa que nenhum valor particular tem preferência.

5.3.3.2 Priori de Jeffreys

Uma conhecida priori não-informativa, que representa uma situação com pouca informação a priori sobre os parâmetros, foi proposta por Jeffreys (1967). Seja X uma única observação com função de (densidade) de probabilidade $p(\mathbf{x}|\boldsymbol{\theta})$. A medida de informação esperada de Fisher de $\boldsymbol{\theta}$ através de X é definida como,

$$I_F(\boldsymbol{\theta}) = \mathbb{E} \left[-\frac{\partial^2 \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right].$$

Assim, a distribuição a priori proposta por Jeffreys é dada por:

$$h(\boldsymbol{\theta}) \propto |I_F(\boldsymbol{\theta})|^{1/2}.$$

Vale resaltar, que a priori de Jeffreys viola o princípio da verossimilhança, já que a informação de Fisher depende da distribuição amostral. Mas um importante ganho desta distribuição a priori é o fato de que o parâmetro $\boldsymbol{\theta}$ não precisa necessariamente pertencer a um espaço limitado.

Bernardo (1989) chama a atenção para alguns aspectos da distribuição a priori de Jeffreys:

- A principal motivação intuitiva da distribuição a priori de Jeffreys é ser invariante a reparametrizações, a qual é uma condição necessária mas não suficiente, para determinar uma referência sensível de distribuição a priori.
- A existência da distribuição a priori de Jeffreys requer condições de regularidade fortes, como a normalidade assintótica da distribuição a posteriori de $\boldsymbol{\theta}$, conforme Bernardo e Smith (1994).

Detalhes sobre outras distribuições a priori não-informativas podem ser obtidos, por exemplo, em Box e Tiao (1973), Berger (1985), Bernardo e Smith (1994), dentre outros autores.

5.4 Metodologia inferencial

5.4.1 Predição pontual

A própria ideia de predição de parâmetros por pontos conduz no cenário bayesiano a tomar como preditores pontos típicos da distribuição a posteriori, para o que se revela útil a determinação das suas medidas de localização.

A escolha das estimativas bayesianas de θ depende naturalmente da forma de $p(\theta|\mathbf{x})$, bem como dos objetivos do seu uso. As estimativas mais usadas são: a média a posteriori, a moda a posteriori e a mediana a posteriori (de parâmetros escalares), cuja a definição se apresenta a seguir (Paulino et. al, 2003):

- Média a posteriori

$$\hat{\theta} = \mathbb{E}(\theta|\mathbf{x}) = \int_{\Theta} \theta p(\theta|\mathbf{x}) d\theta.$$

- Mediana a posteriori

$$\mathbb{P}\{\theta \geq \hat{\theta}|\mathbf{x}\} \geq \frac{1}{2} \quad \text{e} \quad \mathbb{P}\{\theta \leq \hat{\theta}|\mathbf{x}\} \geq \frac{1}{2}.$$

- Moda a posteriori

$$\hat{\theta} = \text{Moda}(\theta|\mathbf{x}) = \max_{\theta \in \Theta} p(\theta|\mathbf{x}) = \max_{\theta \in \Theta} L(\theta|\mathbf{x})p(\theta).$$

5.4.2 Funções de perda

Dentre as funções de perda mais comuns, destaca-se as seguintes:

1. Perda quadrática:

$$L(d, \theta) = (d - \theta)^2.$$

Encontrar um preditor bayesiano para θ , consiste em tomar a decisão d sob a penalidade $L(d, \theta)$. Então, minimizando $\mathbb{E}[L(d, \theta)]$, tem-se

$$\begin{aligned}\mathbb{E}[L(d, \theta)] &= \mathbb{E}(d - \theta)^2 \\ &= \mathbb{E}(d^2 - 2d\theta + \theta^2) \\ &= d^2 - 2d\mathbb{E}(\theta) + \mathbb{E}(\theta^2) \\ &= \underbrace{(d - \mathbb{E}(\theta))^2}_{\geq 0} + \underbrace{\text{Var}(\theta)}_{> 0}.\end{aligned}$$

Então para $d = \mathbb{E}(\theta)$, $\mathbb{E}[L(d, \theta)]$ é mínima, logo $d = \mathbb{E}(\theta)$ é a decisão ótima quando a penalidade é perda quadrática.

2. Perda absoluta:

$$L(d, \theta) = a|d - \theta|, a > 0.$$

É possível mostrar que sob perda absoluta o preditor bayesiano é a mediana a posteriori, $\hat{\theta} = \text{Med}(\theta, x)$.

3. Perda zero-um:

$$\begin{cases} 0 & \text{se, } |d - \theta| \leq b, b \text{ é constante} \\ 1 & \text{se, } |d - \theta| > b \end{cases}$$

Aqui podemos mostrar que sob perda zero-um o preditor bayesiano será a moda a posteriori, $\hat{\theta} = \text{Moda}(\theta, x)$.

5.4.3 Predição Intervalar

Sob o ponto de vista bayesiano, a forma mais adequada de avaliar a informação disponível a respeito de parâmetros desconhecidos é por meio da distribuição a posteriori (Migon e Gamerman, 1999). Ao resumir a distribuição a posteriori em um único valor, não se tem uma medida da precisão da estimativa obtida. Uma alternativa para contornar esta situação é obter intervalos de credibilidade para estes valores.

A definição de intervalo de credibilidade bayesiano é descrita a seguir:

Uma região $C \subset \Theta$ é um intervalo de credibilidade bayesiano $100(1 - \alpha)\%$ para θ se $p(\theta \in C|\mathbf{x}) \geq 1 - \alpha$, em que $1 - \alpha$ é o nível de credibilidade.

Um tipo especial de intervalo de credibilidade, conhecido como HPD (“Highest Posterior Density”), é construído de tal forma que a densidade de probabilidade dos pontos dentro do intervalo é maior que a de todos os pontos fora dele e a probabilidade do verdadeiro valor do parâmetro estar entre os limites definidos pelo intervalo é igual a $1 - \alpha$, onde $0 < \alpha < 1$ é escolhido arbitrariamente.

5.5 Abordagem bayesiana para modelos lineares mistos

Para se fazer uma análise estatística através das técnicas bayesianas para o modelo linear misto normal especificado em (2.3), primeiramente precisa-se definir a distribuição a priori conjunta para os parâmetros desconhecidos, que são $\boldsymbol{\beta}$, \mathbf{D} e $\boldsymbol{\psi}$. Ou seja, precisa-se determinar a priori conjunta $p(\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\psi})$. Usualmente, supõe-se independência a priori entre estes parâmetros, tal que $p(\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\psi}) = p(\boldsymbol{\beta})p(\mathbf{D})p(\boldsymbol{\psi})$. Assumindo independência a priori, necessita-se especificar distribuições para os parâmetros $\boldsymbol{\beta}$, \mathbf{D} e $\boldsymbol{\psi}$ separadamente. Note que o vetor de efeitos aleatórios \mathbf{b} já tem sua distribuição de probabilidade especificada pelas suposições do modelo.

Segundo a formulação do modelo misto visto como um modelo hierárquico definida nas equações (2.7) e (2.8), as distribuições a priori $p(\boldsymbol{\beta})$, $p(\mathbf{D})$, $p(\boldsymbol{\psi})$ correspondem a um terceiro estágio da hierarquia.

A distribuição a priori conjugada usual para $\boldsymbol{\beta}$ é

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}^*, \mathbf{G}),$$

em que $\boldsymbol{\beta}^*$ é um vetor conhecido de dimensão $p \times 1$ e \mathbf{G} é uma matriz de variâncias e covariâncias conhecida, de dimensão $p \times p$. Segundo Wu e Zhang (2006), a matriz \mathbf{G} é flexível. Por exemplo, $\mathbf{G} = \lambda \mathbf{I}_p$, indica que os componentes de $\boldsymbol{\beta}$ são independentes. Além disso, quando $\lambda \rightarrow \infty$, tem-se que $\mathbf{G}^{-1} = \lambda^{-1} \mathbf{I}_p \rightarrow 0$. Isso indica que o limite a priori de $\boldsymbol{\beta}$ é não informativo.

As distribuições a priori para \mathbf{D} e $\boldsymbol{\psi}$ dependerão da estrutura de covariância adotada para cada uma delas. Em modelos normais, o mais usual é parametrizar componentes de variância e matrizes de variâncias e covariâncias como componentes e matrizes de precisão, pois isto facilita os cálculos das distribuições a posteriori. Um componente de precisão é o recíproco do componente de variância correspondente, e uma matriz de precisão é a

inversa da correspondente matriz de variância.

Se a estrutura de covariância VC (componentes de variância) é adotada para a distribuição a priori para os componentes e matrizes de precisão, então a distribuição Gama pode ser usada como priori para cada um dos componentes de precisão da matriz, ou seja, $\sigma_w^{-2} \sim \text{Gama}(a_w, b_w)$. A priori não-informativa neste caso é obtida quando $a_w, b_w \rightarrow 0$. Outra distribuição a priori não-informativa que poderia ser obtida especificando $\sigma_w \sim \mathcal{U}(0, c_w)$, em que $c_w \rightarrow 0$. Neste caso, a distribuição de probabilidade para σ_w^{-2} tem densidade dada por

$$f(\sigma_w^{-2}) = \frac{(\sigma_w^{-2})^{-3/2}}{2c_w}, c_w^{-2} \leq \sigma_w^{-2} < \infty.$$

Esta última tem sido considerada uma opção melhor que a priori Gama principalmente para componentes de variância da matriz \mathbf{D} .

Se uma estrutura AR(1) é usada para a matriz de variâncias e covariâncias, pode-se especificar uma distribuição Gama para o componente de precisão σ^{-2} e uma priori para o parâmetro ρ poderia ser $\rho \sim \mathcal{U}(a, b)$, em que $a < b$ e $a \geq -1$ e $b \leq 1$. Se $a = -1$ e $b = 1$ são especificados, então se obtém-se uma distribuição a priori não-informativa para ρ .

No caso da estrutura NE (não estruturada), não é necessário assumir uma priori para cada componente de variância da matriz. As distribuições Wishart e Inversa-Wishart são distribuições de probabilidade para, respectivamente, as matrizes de precisão e de variâncias e covariâncias. Além disso, são prioris conjugadas para, respectivamente, a matriz de precisão e de variâncias e covariâncias de uma distribuição normal multivariada. Segundo Gelman et al. (1997), se $\mathbf{W} \sim \text{Wishart}(\rho, \mathbf{S})$, em que \mathbf{W} é uma matriz positiva definida simétrica $k \times k$, ρ é o número de graus de liberdade e \mathbf{S} é uma matriz de escala positiva-definida simétrica $k \times k$, então a função densidade de probabilidade é dada por

$$f_{\mathbf{W}}(\mathbf{w}) = \left[2^{\rho k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\rho+1-i}{2}\right) \right]^{-1} |\mathbf{S}|^{-\rho/2} |\mathbf{w}|^{\frac{(\rho+k+1)}{2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{S}^{-1}\mathbf{w})\right].$$

Da mesma forma, se $\mathbf{W} \sim \text{Inv-Wishart}(\rho, \mathbf{S}^{-1})$, então a densidade é dada pela seguinte expressão:

$$f_{\mathbf{W}}(\mathbf{w}) = \left[2^{\rho k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\rho+1-i}{2}\right) \right]^{-1} |\mathbf{S}|^{\rho/2} |\mathbf{w}|^{\frac{-(\rho+k+1)}{2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{S}\mathbf{w}^{-1})\right].$$

A esperança da distribuição Wishart é $\rho\mathbf{S}$, e da Inversa-Wishart é $(\rho - k - 1)^{-1}\mathbf{S}$. Por isso, nas aplicações bayesianas, quando se deseja atribuir uma priori Wishart para uma

matriz de precisão, geralmente é parametrizada como Wishart $\{\rho, (\rho \mathbf{D}^*)^{-1}\}$, em que \mathbf{D}^* é uma matriz $k \times k$. Assim, \mathbf{D}^* será a média a priori da matriz de variâncias que se deseja estimar, e seus valores podem ser escolhidos conforme as expectativas para os possíveis valores da matriz de variâncias. Uma priori não-informativa tanto para a distribuição Wishart quanto para a Inversa-Wishart é obtida quando $\rho = k$.

Após a especificação das distribuições a priori apropriadas para os parâmetros $\boldsymbol{\beta}$, \mathbf{D} e $\boldsymbol{\psi}$, levando em consideração as estruturas de covariância assumidas para as matrizes $\boldsymbol{\beta}$ e \mathbf{D} , o próximo passo seria a derivação da distribuição a posteriori. Porém, independente da estrutura escolhida, não é possível determinar analiticamente a distribuição a posteriori, ou reconhecer o núcleo de alguma distribuição multivariada conhecida através do caminho da proporcionalidade. O algoritmo Amostrador de Gibbs (Gamerman, 1997; Robert e Casella, 1999) pode ser aplicado neste caso, já que, para os modelos mistos, é possível derivar as distribuições condicionais completas.

5.5.1 Relação entre o BLUE e o BLUP e os estimadores bayesianos

Segundo Wu e Zhang (2006), os melhores estimadores e preditores lineares não viciados que maximizam (2.13), no contexto clássico, são iguais ao limite das esperanças a posteriori quando $\mathbf{G}^{-1} \rightarrow \mathbf{0}$. Isto é:

$$\hat{\boldsymbol{\beta}} = \lim_{\mathbf{G}^{-1} \rightarrow \mathbf{0}} \mathbb{E}(\boldsymbol{\beta}|\mathbf{y}) \quad \text{e} \quad \hat{\mathbf{b}} = \lim_{\mathbf{G}^{-1} \rightarrow \mathbf{0}} \mathbb{E}(\mathbf{b}|\mathbf{y})$$

Além disso, como $\mathbf{G}^{-1} \rightarrow \mathbf{0}$, as distribuições a posteriori são dadas pelas seguintes expressões:

$$\begin{aligned} \boldsymbol{\beta}|\mathbf{y} &\sim \mathcal{N}[\hat{\boldsymbol{\beta}}, (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}], \\ \mathbf{b}|\mathbf{y} &\sim \mathcal{N}(\hat{\mathbf{b}}, \mathbf{D} - \mathbf{DZ}^\top \mathbf{P}_V \mathbf{DZ}), \\ \mathbf{e}|\mathbf{y} &\sim \mathcal{N}(\hat{\mathbf{e}}, \boldsymbol{\psi} - \boldsymbol{\psi} \mathbf{P}_V \boldsymbol{\psi}), \end{aligned} \tag{5.1}$$

em que $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}$ e

$$\mathbf{P}_V = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}.$$

Note que $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{b}}$ envolvem as matrizes \mathbf{D} e $\boldsymbol{\psi}$ que geralmente são desconhecidas. Substituindo essas matrizes pelos seus respectivos estimadores, os estimadores bayesianos, $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{b}}$ são usualmente conhecidas como estimadores empíricos bayesianos, embora a es-

timação empírica bayesiana é convencionalmente aplicada somente ao efeito aleatório \mathbf{b}_i , $i = 1, \dots, q$.

Como já falado anteriormente, as expressões (4.1) fornecem os limites a posteriori de $\boldsymbol{\beta}$, \mathbf{b} e \mathbf{e} , sob o enfoque bayesiano, quando $\mathbf{G}^{-1} \rightarrow \mathbf{0}$ ou quando a priori de $\boldsymbol{\beta}$ é não negativa. Às vezes, é interessante conhecer as distribuições a posteriori de \mathbf{b} e \mathbf{e} quando $\boldsymbol{\beta}$ é fornecido, isto é, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. Assim, como descrito em Wu e Zhang (2006), as distribuições a posteriori de interesse são dadas pelas seguintes expressões:

$$\begin{aligned}\mathbf{b}|\mathbf{y}, \boldsymbol{\beta} &\sim \mathcal{N}[\mathbf{DZ}^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \mathbf{D} - \mathbf{DZ}^\top \mathbf{V}^{-1} \mathbf{ZD}], \\ \mathbf{e}|\mathbf{y}, \boldsymbol{\beta} &\sim \mathcal{N}[\boldsymbol{\psi} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \boldsymbol{\psi} - \boldsymbol{\psi} \mathbf{V}^{-1} \boldsymbol{\psi}].\end{aligned}$$

Uma ideia semelhante da extensão do BLUP dentro do contexto bayesiano pode ser encontrada em Lindley e Smith (1972); Dempfle (1977).

Segundo Robinson (1991), dentro da ideologia bayesiana, há pouca razão para se fazer distinção entre os efeitos fixos e e os efeitos aleatórios. Todos os efeitos são tratados como aleatórios no sentido de que as distribuições de probabilidade utilizadas para descrever a incerteza não são tratadas de forma diferente das distribuições de probabilidade utilizadas para descrever a variação.

6 APLICAÇÃO

No presente capítulo, serão apresentadas as análises estatísticas referentes ao estudo já citado anteriormente, o estudo realizado pela Faculdade de Odontologia da Universidade de São Paulo, para comparar dois tipos de escova de dentes (monobloco e convencional) com relação a eficácia na remoção de placa bacteriana, cujo o conjunto de dados encontra-se na Tabela 2.

6.1 Índice de placa bacteriana

De acordo com Singer e Andrade (1997), que analisaram um conjunto diferente de dados proveniente do mesmo estudo, o modelo proposto é dado da seguinte forma:

$$y = \beta x^\delta \xi, \quad (6.1)$$

em que x é o índice de placa bacteriana pré-escovação, y é o índice de placa bacteriana pós-escovação, $\beta > 0$ é um coeficiente de placa bacteriana residual, δ é um coeficiente de uniformidade da taxa de placa bacteriana residual esperada e ξ um erro aleatório não-negativo.

Para os autores, modelos dessa natureza devem apresentar as seguintes características:

1. Um índice pré-tratamento nulo implica um índice pós-tratamento também nulo;
2. Os índices pré-tratamento e pós-tratamento são não-negativos;
3. Os dados são possivelmente heteroscedásticos (pois são não-negativos e satisfazem a desigualdade $y \leq x$);

4. A relação entre os índices pré-tratamento e pós-tratamento é possivelmente não-linear;
5. As observações realizadas numa mesma unidade experimental são possivelmente correlacionadas.

Com base no que foi proposto por Sef (1999), Nobre (2004) propôs o seguinte modelo para os dados apresentados na Tabela 2:

$$y_{ijk} = \beta_{jk} x_{ijk}^{\delta_{jk}} \xi_{ijk}, \quad (6.2)$$

com $\beta_{jk} > 0$, $i = 1, \dots, 32$, $j = 0, 1$ e $k = 1, 2, 3, 4$, em que $y_{ijk}(x_{ijk})$ são os índices de placa bacteriana pós-teste e pré-teste, respectivamente, para o i -ésimo indivíduo avaliado na k -ésima sessão com a j -ésima escova, β_{jk} é o efeito fixo associado ao j -ésimo tipo de escova na k -ésima sessão, δ_{jk} é o coeficiente de uniformidade do índice de redução de placa bacteriana associado ao j -ésimo tipo de escova na k -ésima sessão, ξ_{ijk} é um erro aleatório não-negativo relativo à i -ésima criança com a j -ésima escova na k -ésima sessão de avaliação.

Claramente o modelo multiplicativo acima é não-linear, mas pode ser linearizado por uma transformação logarítmica, o que resulta em:

$$\ln(y_{ijk}) = \ln(\beta_{jk}) + \delta_{jk} \ln(x_{ijk}) + \ln(\xi_{ijk}),$$

podendo ser reparametrizado por

$$y_{ijk}^* = \lambda_{jk} + \delta_{jk} x_{ijk}^* + \xi_{ijk}^*,$$

em que $y_{ijk}^* = \ln(y_{ijk})$, $\lambda_{jk} = \ln(\beta_{jk})$, $x_{ijk}^* = \ln(x_{ijk})$. Assumimos que $\ln(\xi_{ijk}) = \xi_{ijk}^*$ tem distribuição normal com parâmetros a serem especificados. Para satisfazer a característica (5), vamos considerar

$$\xi_{ijk}^* = \varphi_i + \varepsilon_{ijk},$$

com $\varphi_i \sim \mathcal{N}(0, \tau^2)$ e $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$, denotando respectivamente, o efeito aleatório da i -ésima criança e o erro de medida.

Assim como citado em Nobre (2004) e Nobre e Singer (2007), a seguir descrevemos a estratégia da análise utilizada para simplificar o modelo saturado (5.2):

1. Testar a homogeneidade entre os coeficientes de uniformidade para as duas escovas

nas quatro sessões de avaliação ($\delta_{jk} = \delta, j = 0, 1$ e $k = 1, \dots, 4$), ou seja, reduzir o modelo (5.2) para:

$$y_{ijk} = \beta_{jk} x_{ijk}^{\delta} \xi_{ijk}; \quad (6.3)$$

2. Testar a significância da interação e dos efeitos principais dos tipos de escova com relação aos coeficientes de placa bacteriana residual, ou seja,

$$\beta_{01}/\beta_{11} = \beta_{02}/\beta_{12} = \beta_{03}/\beta_{13} = \beta_{04}/\beta_{14},$$

ou equivalentemente

$$\lambda_{01} - \lambda_{11} = \lambda_{02} - \lambda_{12} = \lambda_{03} - \lambda_{13} = \lambda_{04} - \lambda_{14},$$

e

$$\beta_{jk} = \beta_j,$$

ou equivalentemente,

$$\lambda_{jk} = \lambda_j,$$

3. Ajustar o modelo que incorpora as conclusões obtidas em (1) e (2), ou seja, reduzir o modelo para:

$$y_{ijk} = \beta_j x_{ijk}^{\delta} \xi_{ijk}. \quad (6.4)$$

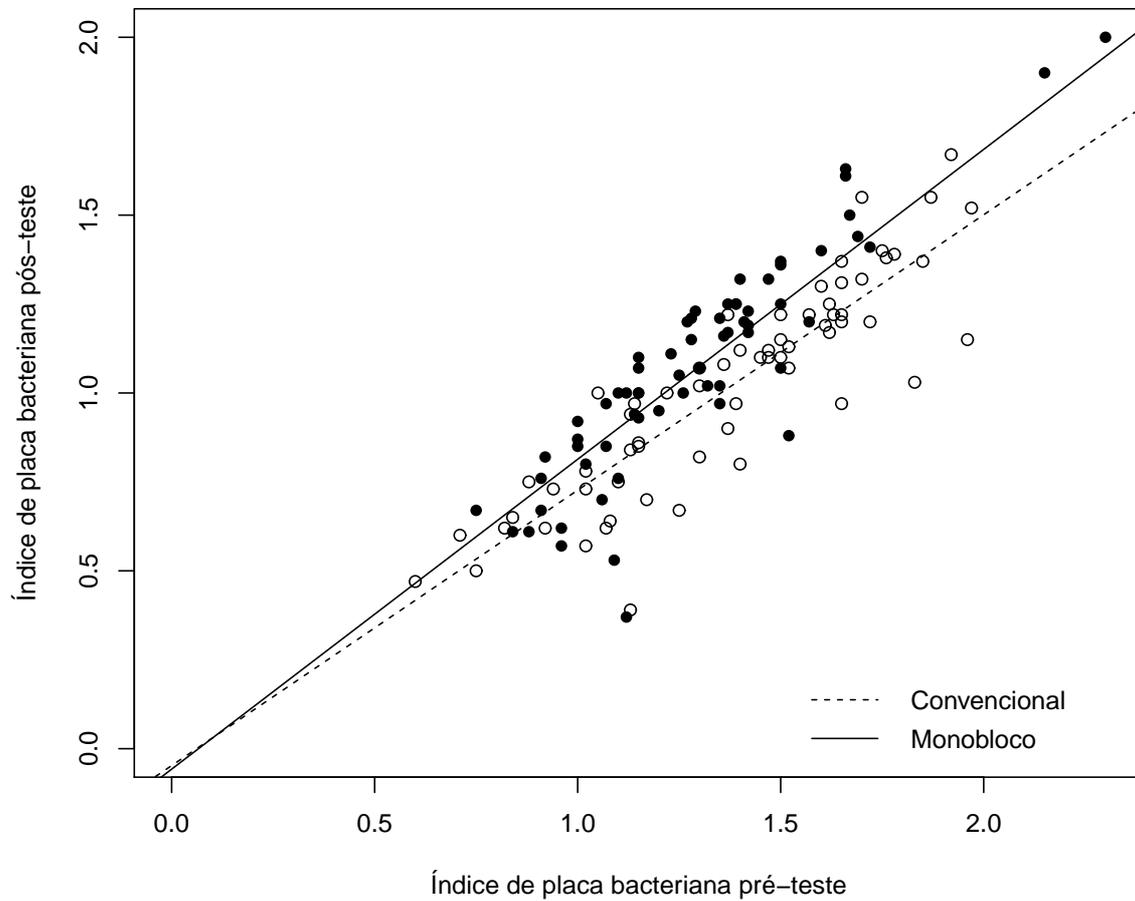


Figura 5: Ajuste do modelo final

As observações representadas por \bullet referem-se às crianças que utilizaram escova do tipo monobloco.

Baseando-se em todas essas considerações, usaremos o modelo assim para ilustrar os procedimentos de diagnóstico. A Figura 6 (a), nos permite checar a linearidade dos efeitos. Nesse caso, há indícios da veracidade da hipótese de linearidade, pois não se observa nenhum tipo de tendência do resíduo marginal conforme o valor de $\ln(x_{ijk})$. Na Figura 6 (b), podemos perceber que a unidade experimental #29 apresenta um comportamento diferenciado das demais unidades experimentais. Na Figura 7 apresenta-se o gráfico de envelope simulado que não fornece indícios de afastamento da normalidade por parte do resíduo com confundimento mínimo. No gráfico que diz respeito ao resíduo condicional padronizado *versus* as unidades experimentais, podemos destacar a observação #12.2, que diz respeito a segunda sessão da criança #12 e a observação #29.4, que diz respeito a quarta sessão da criança #29. Na Figura 8, o gráfico com a distância de Mahalanobis,

também destacamos a criança #29.

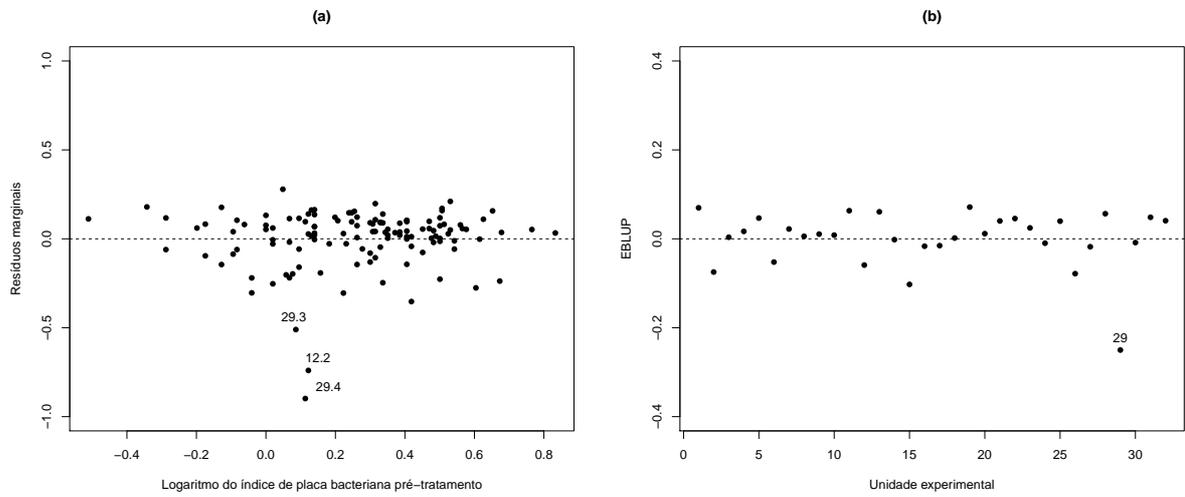


Figura 6: (a) Resíduo marginal (b) EBLUP

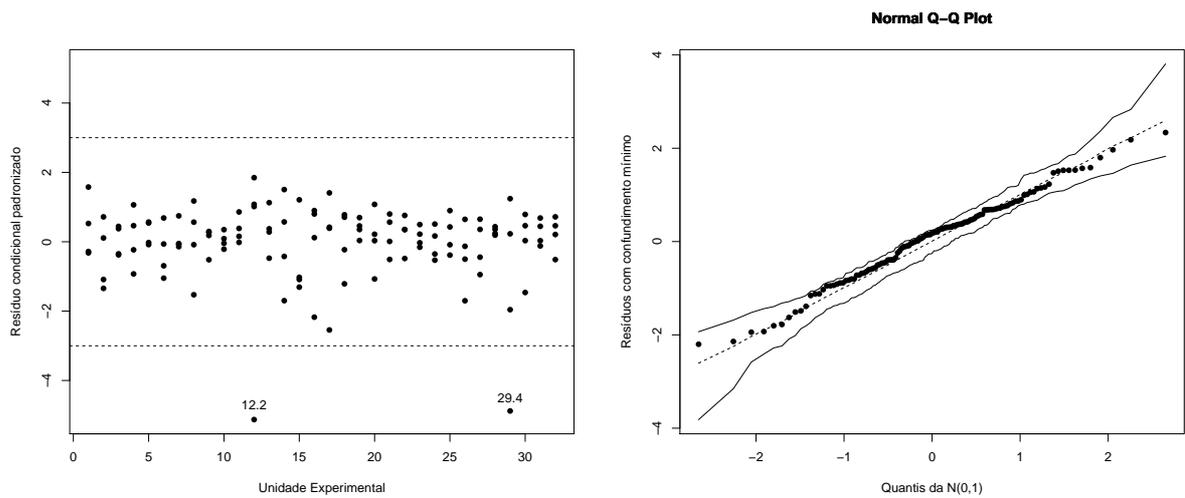


Figura 7: Resíduo condicional padronizado e envelope simulado com 95% para o resíduo com confundimento mínimo.

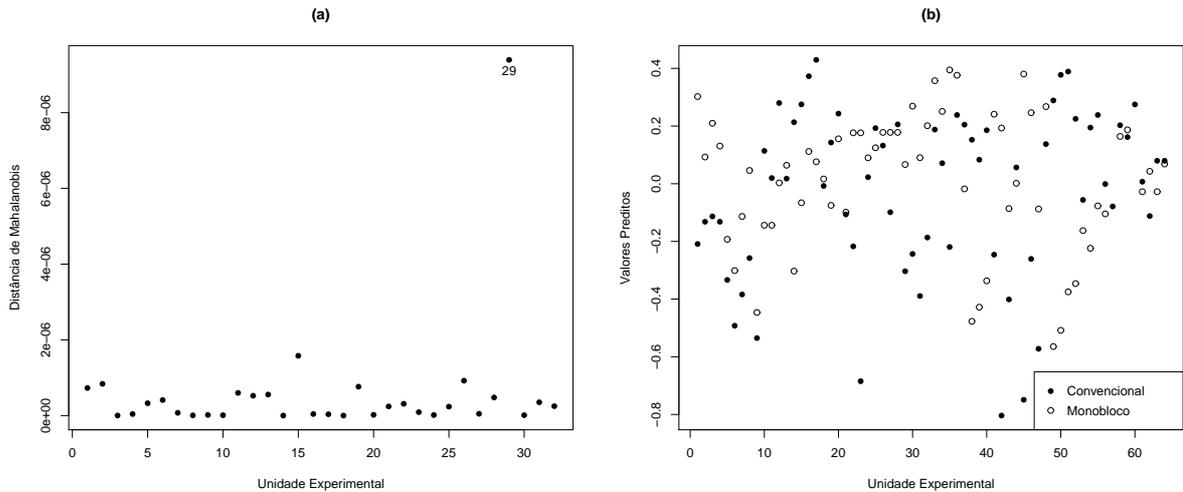


Figura 8: (a)Distância de Mahalanobis (b)Valores preditos para cada tipo de escova

Assim como apresentado em Nobre (2004), a seguir serão apresentadas as estimativas dos parâmetros, bem como seus intervalos de confiança a um nível de confiança de 95% para o modelo completo e para o modelo quando retiramos as unidades experimentais #12 e #29 individualmente e simultaneamente.

Tabela 5: Estimativas dos parâmetros e intervalos de confiança a um nível de confiança de 95% ao eliminar as unidades experimentais #12 e #29.

Parâmetros	β_0	β_1	δ	τ^2	σ^2
Modelo completo	0,72 [0,70; 0,74]	0,81 [0,79; 0,82]	1,06 [0,96; 1,14]	0,007 [0,006; 0,022]	0,021 [0,016; 0,027]
Excluindo #12	0,72 [0,70; 0,74] (0,00%)	0,80 [0,79; 0,82] (-1,23%)	1,06 [0,96; 1,15] (0,00%)	0,007 [0,006; 0,011] (16,67%)	0,015 [0,012; 0,020] (-28,57%)
Excluindo #29	0,72 [0,69; 0,74] (0,00%)	0,83 [0,81; 0,84] (2,47%)	1,07 [0,97; 1,16] (0,94%)	0,001 [0,001; 0,025] (-83,33%)	0,017 [0,015; 0,025] (-19,05%)
Excluindo #12 e #29	0,72 [0,70; 0,74] (0,00%)	0,83 [0,81; 0,84] (2,47%)	1,07 [0,95; 1,15] (0,94%)	0,003 [0,002; 0,041] (-50,00%)	0,012 [0,009; 0,015] (-42,86%)

Segundo Nobre (2004) e Nobre e Singer (2007), conclui-se que as crianças #12 e #29 são influentes no que diz respeito aos parâmetros de covariância, em particular, a criança #12 é influente no que diz respeito a σ^2 e a criança #29 com respeito à τ^2 . Tais resultados são compatíveis com as análises residuais. Uma descrição das características dos indivíduos influentes segue:

#12: Essa criança utilizou a escova convencional e apresentou o segundo menor índice de placa bacteriana pré-escovação (0,71) e pós-escovação (0,39), ambas na segunda sessão; apresenta também um alto índice de placa bacteriana pós-escovação, entre as 25% contradizendo o modelo ajustado, que prevê maiores índices de placa bacteriana pós-escovação para crianças que utilizam a escova monobloco; essa criança também apresenta a maior variabilidade entre os índices de placa bacteriana pós-escovação e a segunda maior variância entre os valores de índice de placa bacteriana pré-escovação para as quatro sessões de

avaliação além de apresentar o segundo menor índice de redução de placa bacteriana (y/x);

#29: Essa criança apesar de ter utilizado a escova monobloco, apresentou todos seus índices de placa bacteriana pós-escovação entre os 25% menores índices, inclusive o menor (0,37) obtido na quarta sessão; este resultado contraria o esperado sob o modelo ajustado, que prevê menores índices para crianças que utilizaram a escova convencional; apresentou também dois entre os três menores, incluindo o menor, índices de redução de placa bacteriana (y/x).

Assim, no exemplo considerado, a influência esperada na retirada de uma unidade experimental é de $(1/32) \times 100 = \pm 3,13\%$. Quando eliminamos a unidade experimental #29 a estimativa de τ^2 decresce 83,33%, implicando uma alta influência da respectiva unidade experimental na estimativa da variância do efeito aleatório da criança. Toda inferência realizada com base na amostra completa, continua válida, indicando que o modelo ajustado é robusto.

A partir da análise do gráfico do modelo ajustado e da Tabela 5 referente a estimativa dos parâmetros do modelo final, é razoável concluir que a escova convencional mostrou-se mais eficaz na redução do índice de placa bacteriana, $\hat{\beta}_0 < \hat{\beta}_1$.

A seguir, será apresentada uma análise bayesiana do estudo em questão. A fim de viabilizar a análise e facilitar a obtenção dos resultados desejados, fez-se uso do software *OpenBUGS* (*Bayesian inference Using Gibbs Sampling*).

Inicialmente, linearizando por meio de uma transformação logarítmica, o modelo (5.4) pode ser escrito da seguinte forma:

$$\ln(y_{ijk}) = \ln(\beta_j) + \delta \ln(x_{ijk}) + \ln(\xi_{ijk}),$$

podendo ser reparametrizado

$$y_{ijk}^{**} = \lambda_j + \delta x_{ijk}^{**} + \xi_{ijk}^{**},$$

em que $y_{ijk}^{**} = \ln(y_{ijk})$, $\lambda_j = \ln(\beta_j)$, $x_{ijk}^{**} = \ln(x_{ijk})$ e $\ln(\xi_{ijk}) = \xi_{ijk}^{**}$.

Assim como proposto anteriormente , tem-se que:

$$\begin{aligned}\xi_{ijk}^{**} &= \varphi_i + \varepsilon_{ijk} \\ \varphi_i &\sim \mathcal{N}(0, \tau^2) \\ \varepsilon_{ijk} &\sim \mathcal{N}(0, \sigma^2)\end{aligned}$$

Dessa forma,

$$\begin{aligned}y_{ijk}^{**} &\sim \mathcal{N}(\mu_{ijk}, \zeta) \\ y_{ijk}^{**} | \lambda_j, \delta, \tau^2, \sigma^2 &\sim \mathcal{N}(\mu_{ijk}, \zeta) \\ \zeta = \text{Var}(y_{ijk}^{**} | \lambda_j, \delta, \tau^2, \sigma^2) &= \text{Var}(\xi_{ijk}^{**}) \\ &= \text{Var}(\varphi_i + \varepsilon_{ijk}) \\ &\stackrel{\text{ind}}{=} \tau^2 + \sigma^2\end{aligned}$$

Em síntese, tem-se que:

$$\left\{ \begin{array}{l} y_{ijk}^{**} \sim \mathcal{N}(\mu_{ijk}, \tau^2 + \sigma^2) \\ \lambda_j \sim p(\lambda_j) \\ \delta \sim p(\delta) \\ \tau^2 \sim p(\tau^2) \\ \sigma^2 \sim p(\sigma^2) \end{array} \right.$$

em que $p(\lambda_j)$, $p(\delta)$, $p(\tau^2)$, $p(\sigma^2)$ são as distribuições a priori adotadas para λ_j , δ , τ^2 e σ^2 , respectivamente.

As distribuições a priori aqui propostas são prioris não informativas, visto que a informação que se tem a respeito dos parâmetros do modelo não é significativa se comparada à informação proveniente da amostra.

Foi considerada para λ_j uma distribuição a priori Normal com média zero e variância igual a 1000. Para δ , τ^2 e σ^2 foram adotadas distribuições a priori Gama e Exponencial com variância suficientemente grande, com o objetivo de avaliar a sensibilidade de tais prioris. A Figura 9 apresenta o gráfico com as distribuições a priori Gama e Exponencial.

Na Tabela 6, encontra-se as estimativas, média e mediana a posteriori e intervalos de credibilidade de 95% do modelo final, obtidas via software **OpenBUGS** (*Bayesian inference*

Using Gibbs Sampling), para diferentes distribuições a priori. Para a análise foram geradas 10.000 amostras para cada parâmetro de interesse, sendo as 8.000 primeiras descartadas para evitar algum efeito de valores iniciais.

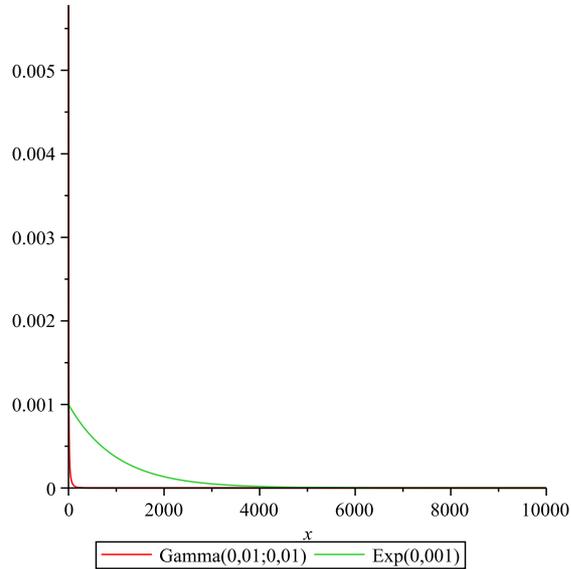


Figura 9: Distribuições a priori

Tabela 6: Estimativas, média e mediana a posteriori e intervalos de credibilidade de 95% do modelo final.

Parâmetros	Prioris	Estimativas		
		Média	Mediana	IC(95%)
β_0	Gama(0,01; 0,01)	0,75	0,75	[0,71; 0,79]
	Exp(0,001)	0,75	0,75	[0,71; 0,79]
β_1	Gama(0,01; 0,01)	0,76	0,76	[0,72; 0,81]
	Exp(0,001)	0,76	0,76	[0,72; 0,80]
δ	Gama(0,01; 0,01)	1,07	1,07	[0,94; 1,19]
	Exp(0,001)	1,07	1,07	[0,94; 1,20]
τ^2	Gama(0,01; 0,01)	0,015	0,009	[0; 0,031]
	Exp(0,001)	0,006	0,002	[0; 0,0259]
σ^2	Gama(0,01; 0,01)	0,016	0,021	[0; 0,042]
	Exp(0,001)	0,025	0,027	[0; 0,037]

Nas Figuras 10 e 11, são apresentadas as distribuições a posteriori e nas Figuras 12 e 13, são apresentadas as trajetórias completas da simulação, para checar a convergência, dos parâmetros do modelo para as diferentes distribuições a priori adotadas.

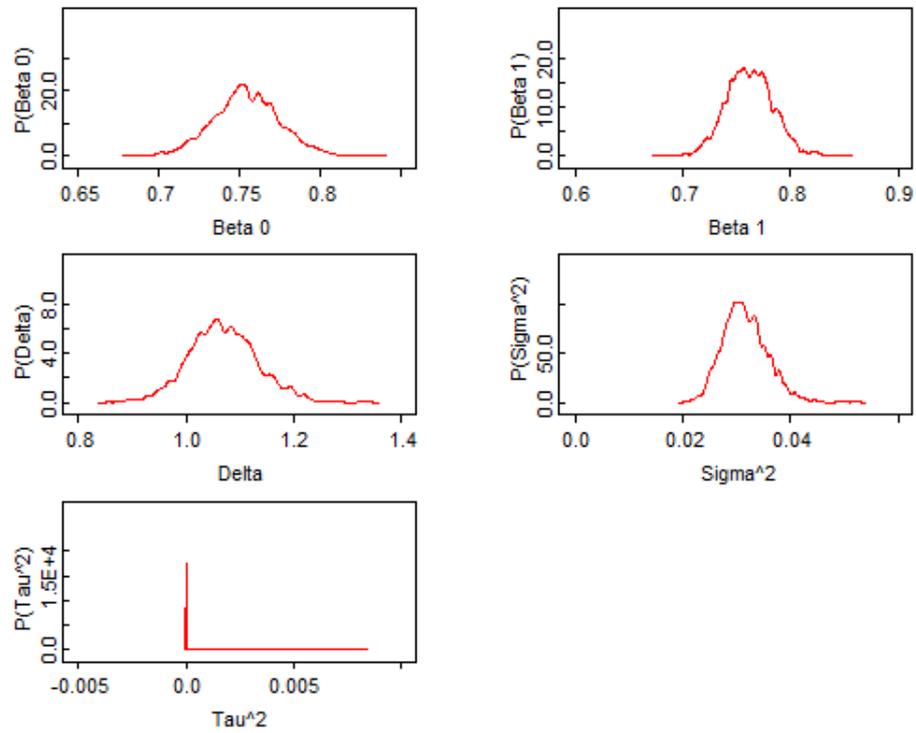


Figura 10: Distribuições a posteriori dos parâmetros com priori $\text{Gama}(0,01; 0,01)$

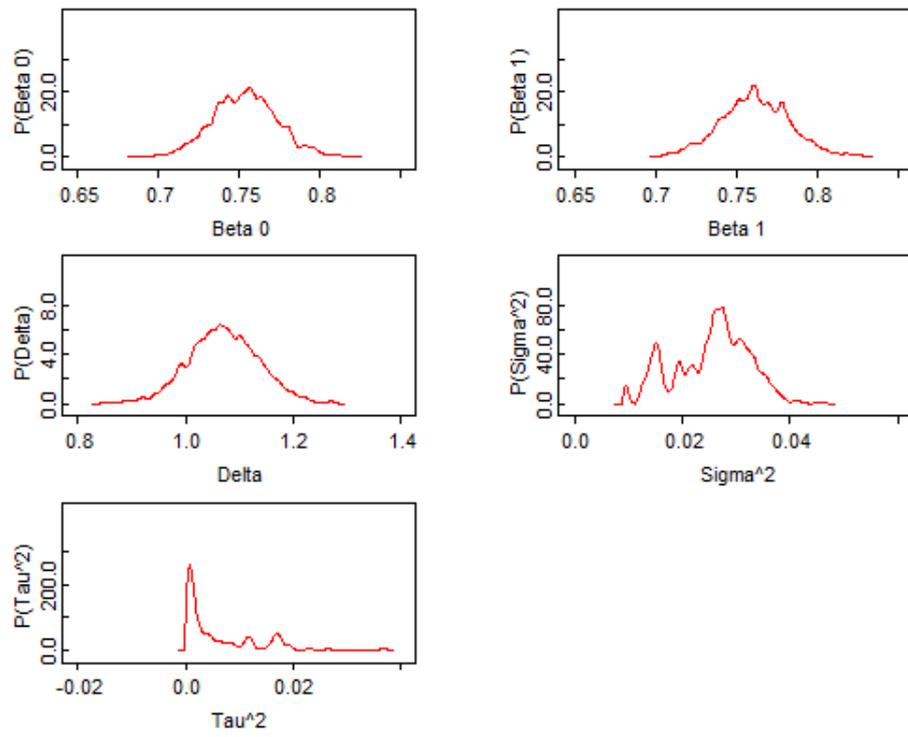


Figura 11: Distribuições a posteriori dos parâmetros com priori Exponencial(0,001)

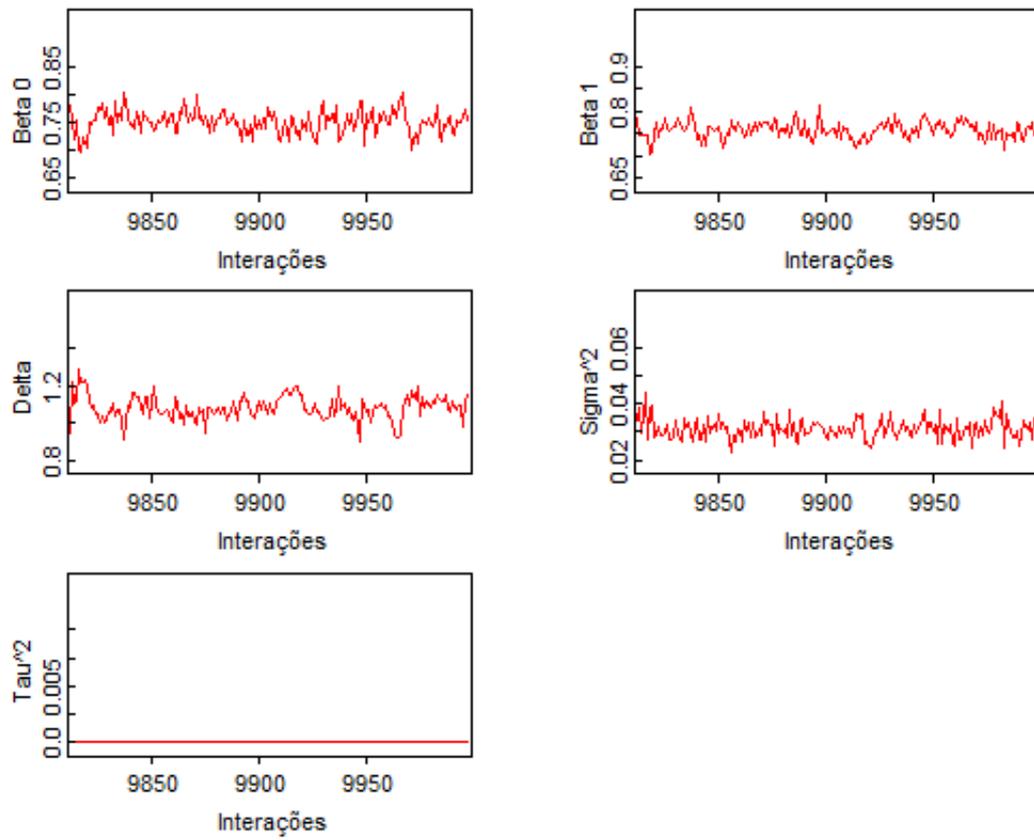


Figura 12: Trajetórias completas da simulação dos parâmetros com priori $\text{Gama}(0,01; 0,01)$

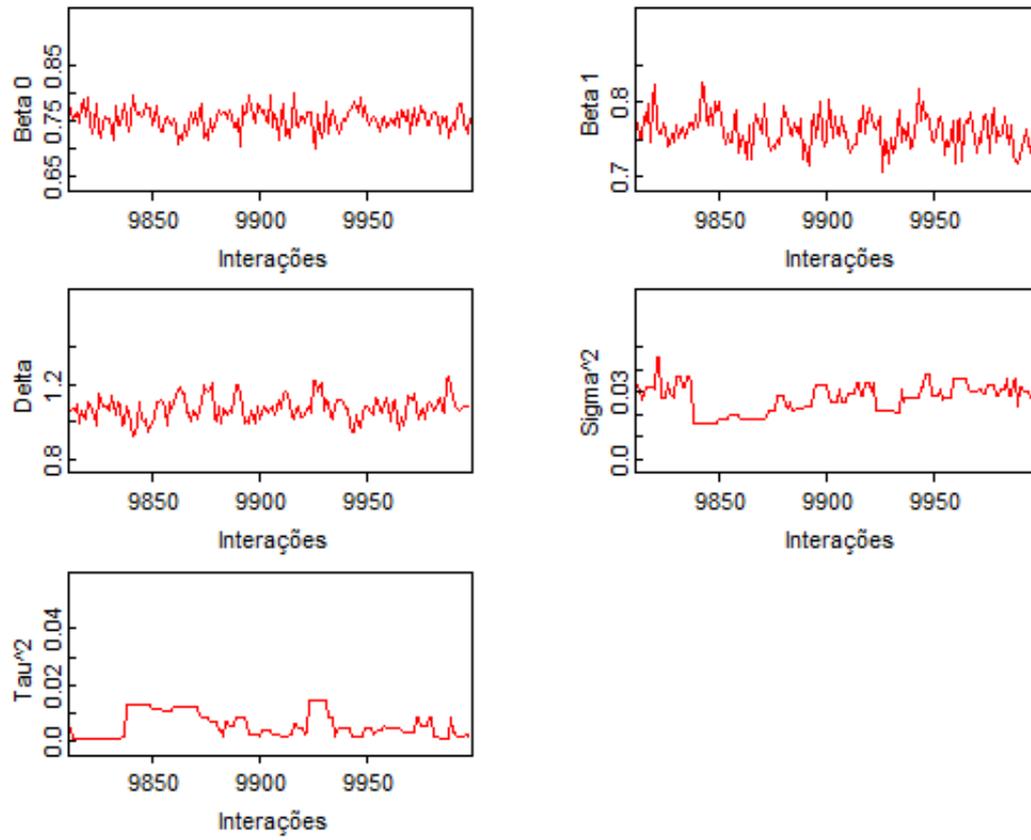


Figura 13: Trajetórias completas da simulação dos parâmetros com priori Exponencial(0,001)

A análise baseada na teoria bayesiana apresentou resultados muito próximos aos obtidos via teoria clássica, dessa forma a conclusão de que a escova convencional apresentou uma maior eficácia na redução do índice de placa bacteriana permanece inalterada.

7 CONCLUSÃO

Os modelos lineares mistos tem sido um tema de crescente interesse em Estatística nos últimos cinquenta anos, pois possibilitam a modelagem de correlação intra indivíduo, muitas vezes presente em dados agrupados. Medições feitas em um mesmo indivíduo não podem ser consideradas não correlacionadas e os modelos lineares mistos constituem uma ferramenta conveniente para modelar essa dependência intra indivíduos.

As análises e interpretações dos dois estudos aqui utilizados (Sef, 1999; Singer et al., 1988) não foram esgotadas ao máximo, é de interesse, em pesquisas futuras, propor uma maior flexibilização dos modelos adotados, bem como a implementação das técnicas de diagnóstico bayesiano para ambos os estudos.

REFERÊNCIAS

1. ANDREONI, S. *Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo*. 1989. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, USP, São Paulo, 1989.
2. AZZALINI A. ; DALLA VALLE A. The multivariate skew-normal distribution. *Biometrika*, v. 83, p. 715-726, 1996.
3. AZZALINI, A. ; CAPITANIO, A. Distributions generated by perturbations of symmetry with emphasis on a multivariate skew t distribution. *Journal of Royal Statistical Society B*, v. 65, p. 367-389, 2003.
4. BANERJEE, M. Cooks distance in linear longitudinal models. *Communications in Statistics, Theory and Methods*, v. 27, p. 2973-2983, 1998.
5. BERGER, J. O. *Statistical decision theory and bayesian analysis*. New York : Springer-Verlag, 1985.
6. BERNARDO, J.M. The Geometry of asymptotic inference: on multivariate Jeffreys prioris. *Statistical Science*, v. 4, p. 227-229, 1989.
7. BERNARDO, J. M. ; SMITH, A. F. M. *Bayesian theory*. New York: Wiley, 1994.
8. BILLOR, N. ; LOYNES, R. M. Local Influence: a new approach. *Commun. Statist. Theory Meth.*, v. 22, p.1595-1661, 1993.
9. BONATE, P. L. *Analysis of pretest-posttest designs*. Boca Raton: Chapman Hall, 2000.
10. BOX, G. E. P. ; TIAO, G. C. *Bayesian inference in statistical analysis*. New York: Wiley Classics, 1973.
11. BOZDOGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, v. 52, p. 345-370, 1987.
12. BROOKS, S. P. Discussion of the paper bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, v. 64, p. 616-618, 2002.

13. CANCHO, V. G. ; ORTEGA, E. M. M. ; PAULA, G. A. On estimation and influence diagnostics for log-Birnbaum Saunders student-t regression models: full bayesian analysis. *Journal of Statistical Planning and Inference*, v. 140, p. 2486-2496, 2010.
14. CARLIN, B. P. ; LOUIS, T. A. *Bayes and empirical Bayes methods for data analysis*. Boca Raton: Chapman and Hall, 2001.
15. CHATTERJEE, S. ; HADI, A. S. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, v. 1, p. 379-393, 1986.
16. CHATTERJEE, S. ; HADI, A. S. *Sensitivity analysis in linear regression*. New York : Wiley, 1988.
17. CHEN, M. H.; SHAO, Q. M. ; IBRAHIM, J. G. *Monte Carlo methods in bayesian computation*. New York: Springer-Verlag, 2000.
18. CHO, H. et al. Bayesian case influence diagnostics for survival models. *Biometrics*, v. 65, p. 116-124, 2009.
19. CHRISTENSEN, R. Exact tests for variance components. *Biometrics*, v. 52, p.309-314, 1996.
20. COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, v. 19, p. 15-18, 1977.
21. COOK, R. D. ; WEISBERG, S. Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, v. 22, p. 495-508, 1980.
22. COOK, R. ; WEISBERG, S. *Residuals and influence in regression*. New York : Chapman and Hall, 1982.
23. COOK, R. D. Assessment of local influence. *Journal of the Royal Statistical Society B*, v. 48, p.133-169, 1986.
24. COSTA, T. R. *Modelos lineares mistos: uma aplicação na produção de leite de vacas da raça Sindi*. 2010. Dissertação (Mestrado em Biometria e Estatística Aplicada) - Universidade Federal Rural de Pernambuco, Recife, 2010.
25. CROWDER, M. J. ; HAND, D. J. *Analysis of repeated measures*. London : Chapman Hall, 1990.
26. CROWDER, M. J. ; HAND, D. J. *Practical longitudinal data analysis*. London : Chapman Hall, 1996.

27. COX, D. R. ; SOLOMON, P. J. *Components of variance*. New York : Chapman Hall, 2003.
28. DAVIS, C. S. *Statistical methods for the analysis of repeated measurements*. New York : Springer, 2002.
29. DEMIDENKO, E. *Mixed models: theory and applications*. New York : Wiley, 2004.
30. DEMIDENKO, E. ; STUKEL, T. A. Influence analysis for linear mixed-effects Models. *Stat. Med.*, v. 24, p. 893-909, 2005.
31. DEMPFFLE, L. Comparison of several sire evaluation methods in dairy cattle breeding. *Livestock Production Science*, v. 4, n. 2, p.129-139, 1977.
32. DIGGLE, P. J. An approach to the analysis of repeated measurements. *Biometrics*, v. 44, p. 959-971, 1988.
33. DIGGLE, P. J.; LIANG, K. ; ZEGER, S. L. *Analysis of longitudinal data*. Oxford : Clarendon Press, 2002.
34. DOGANAKSOY, N. ; BALAKRISHNAN, N. A useful property of best linear unbiased predictors with applications to life-testing. *The American Statistician*, v. 51, p. 22-28, 1997.
35. DRAPER, N. R. ; SMITH, H. *Applied regression analysis*. New York : Wiley, 2002.
36. EFROMOVICH, S. *Nonparametric curve estimation: methods, theory and applications*. New York : Springer, 1999.
37. EHLERS, R. S. *Introdução à inferência bayesiana*. Texto publicado pela Universidade Federal do Paraná. UFPR : Departamento de Estatística, 2007.
38. EUBANK, R. L. *Spline smoothing and nonparametric regression*. New York : Marcel Dekker, 1988.
39. EUBANK, R. L. *Nonparametric regression and spline smoothing*. New York : Marcel Dekker, 1999.
40. FAI, A.H.T. ; CORNELIUS, P.L. Approximate F-Tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal Statistical Computing and Simulation*, v. 54, p. 363-378, 1996.

41. FUNG, W. K. et al. Influence diagnostics and outliers test for semiparametric mixed models. *Journal of the Royal Statistical Society B*, v. 64, p. 565-579, 2002.
42. GAMERMAN D. *Markov chain Monte Carlo*. London: Chapman Hall, 1997.
43. GAMERMAN, D. ; LOPES. H. F. *Markov chain Monte Carlo: stochastic simulation for bayesian inference*. New York: Chapman Hall, 2006.
44. GELMAN, A et al.. *Bayesian data analysis*. London: Chapman Hall, 1997.
45. GIANOLA, D. ; FOULLEY, J. L. Variance estimation from integrated likelihood (VEIL). *Génétique, Sélection, Evolution*, v. 22, p. 403-417, 1990.
46. GRAYBILL, F. A. *Theory and application of the linear model*. Massachusetts: Oux-burg Press, 1976.
47. GREEN, P. J., ; SILVERMAN, B. W. *Nonparametric regression and generalized linear models: a roughness penalty approach*. New York: Chapman Hall, 1994.
48. HARDLE, W. *Applied nonparametric regression*. New York : Cambridge University Press, 1990.
49. HARDLE, W.; LIANG, H. ; GAO, J. *Partially linear models*. Heidelberg: Physica, 2000.
50. HARVILLE, D. A. Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, v. 4, n.2, 237-240, 1976.
51. HARVILLE, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, v. 72, p. 320-340, 1977.
52. HARVILLE, D.A. ; JESKE, D.R. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, v. 87, p. 724-731, 1992.
53. HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, v. 31, p. 423-447, 1975.
54. HENDERSON, C. R. *Applications of linear models in animal breeding*. Guelph: University of Guelph, 1984.

55. HILDEN-MINTON, J. A. *Multilevel diagnostics for mixed and hierarchical linear models*. (PhD Thesis) - University of California, Los Angeles, 1995.
56. IBRAHIM, J., CHEN, M. ; SINHA, D. *Bayesian survival analysis*. New York : Springer Verlag, 2001.
57. JEFFREYS, H. *Theory of probability*. London: Oxford University Press, 1967.
58. JENNRICH, R. I. ; SCHLUCHTER, M. D. Unbalanced repeated measures models with structured covariance matrices. *Biometrics*, v. 42, p. 805-820, 1986.
59. JIANG, J. Reml estimation: asymptotic behavior and related topics. *The Annals of Statistics*, v. 24, p. 255-286, 1996.
60. JIANG, J. A derivation of BLUP - best linear unbiased predictor. *Statistics Probability Letters*, v. 32, p.321-324, 1997.
61. JIANG, J. Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statistica Sinica*, v. 8, p. 861-885, 1998.
62. JONES, R. H. *Longitudinal data with serial correlation: a state-space approach*. London: Chapman Hall, 1993.
63. KACKAR, R.N. e HARVILLE, D.A. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, v. 79, p. 853-862, 1984.
64. LAIRD, N. M. ; WARE, J. M. Random-effects models for longitudinal data. *Biometrics*, v. 38, p. 963-974, 1982.
65. LAIRD, N. M. ; LANGE, N. ; STRAM, D. Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, v. 82, p. 97-105, 1987.
66. LEANDRO, R. A. *Introdução à estatística bayesiana*. Piracicaba, São Paulo : ESALQ/USP 2001.
67. LINDLEY, D. V. ; SMITH, A. F. M. Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B*, v. 34, p.1-41, 1972.
68. LINDSTROM, M. ; BATES, D.M. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, v. 83, p. 1014-1022, 1988.

69. LINDSTROM, M. J. ; BATES, D. M. Nonlinear mixed effects models for repeated measures data. *Biometrics*, v. 46, p. 673-687, 1990.
70. LOBATO JÚNIOR, D. *Influência local em modelos de regressão*. Dissertação (Mestrado em Matemática) - Centro de Ciências e Tecnologia, Universidade Federal de Campina Grande, Paraíba, 2005.
71. LONGFORD, N. T. *Random coefficient models*. New York : Oxford University Press, 1993.
72. MARITZ, J. S. ; LWIN, T. *Empirical Bayes methods*. London : Chapman e Hall, 1989.
73. MCCULLOCH, R. E. Local model influence. *Journal of the American Statistical Association*, v. 84 ,p. 473-478, 1989.
74. MCCULLOCH, C. E. ; SEARLE, S. R. *Generalized linear, and mixed models*. New York : Wiley, 2001.
75. MCCULLOCH, C. E.; SEARLE, S. R.; NEUHAUS, J. M. *Generalized linear, and mixed models*. New York : John Wiley, 2008.
76. MIGON, H.S. ; GAMERMAN, D. *Statistical inference: an integrated approach*. New York :Arnold, 1999.
77. NOBRE, J. S. *Métodos de diagnostico para modelos lineares mistos*. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística, USP, São Paulo, 2004.
78. NOBRE, J. S. ; SINGER, J. M. Residuals analysis for linear mixed models. *Biometrical Journal*, v. 49, p. 863-875, 2007.
79. NOBRE, J.S. ; SINGER, J.M. Leverage analysis for linear mixed models. *Journal of Applied Statistics*, v. 38, p.1063-1072, 2011.
80. OFVERSTEN, F. Exact tests for variance components in unbalanced mixed linear models. *Biometrics*, v. 49, p. 45-57, 1993.
81. OHAGAN, A. *Kendalls advanced theory of statistics: bayesian inference*. New York : Edward Arnold, 1994.

82. 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.
83. PARIZOTTO, S. P. C. O. L. *Estudo comparativo da eficácia da escova dental utilizando-se dois tipos de escova, com ou sem, dentífrício e relação do desgaste das escovas com o controle da placa bacteriana na dentição decídua*. 1999. Dissertação (Mestrado em Odontologia)- FOU SP, São Paulo, 1999.
84. PATTERSON, H. D. ; THOMPSON, R. Recovery of interblock information when block sizes are unequal. *Biometrika*, v. 58, p. 545-554, 1971.
85. PAULA, G. A. *Modelos de regressão com apoio computacional*. São Paulo : IME-USP, 2010.
86. PAULINO, C. D.; TURKMAN, A. A. ; MURTEIRA, B. *Estatística bayesiana*. Lisboa : Fundação Calouste Gulbenkian, 2003
87. PINHEIRO, J. C. ; BATES, D. M. *Mixed-Effects in S and S-PLUS*. New York : Springer, 2000.
88. PINHEIRO, J. C.; LIU, C. ; WU, Y. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate distribution. *Journal of Computational and Graphical Statistics*, v. 10, p. 249-276, 2001.
89. RAO, P. S. *Variance components: mixed models methodologies and applications*. Boca Raton : CRC Press, 1997.
90. RRESENDE, M. D. V. de ; ROSA-PEREZ, J. R. H. *Genética quantitativa e estatística no melhoramento animal*. Curitiba: Imprensa Universitária, Universidade Federal do Paraná, 1999.
91. ROBERT, C.P. ; CASELLA, G. *Monte Carlo statistical methods*. New York : Springer Verlag, 1999.
92. ROBINSON, G. K. The estimation of random effects. *Statistical Science*, v. 6, p. 15-51, 1991.
93. ROCHA, F. M. M. *Seleção de estruturas de covariância para dados com medidas repetidas*. 2004. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística, USP, São Paulo, 2004.

94. RUPPERT, D.; WAND, M. ; CARROLL, R. *Semiparametric regression*. Cambridge : Cambridge University Press, 2003.
95. SAHU, S. K.; DEY, D.E. e BRANCO, M. A new class of multivariate skew distributions with applications to bayesian regression models. *The Canadian Journal of Statistics*, v. 29, p. 217-232, 2003.
96. SAVALLI, C.; PAULA, G. A. ; CYSNEIROS, F. J. A. Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, v. 6, p. 59-76, 2006.
97. SEARLE, S. R. *Matrix algebra useful for statistics*. New York : John Wiley, 1982.
98. SEARLE, S. R.; CASSELA, G. ; MCCULLOGH, C. E. *Variance components*. New York : John Wiley, 1992.
99. SEF, H.C. *Modelos de regressão multiplicativos para dados pré-teste/pós-teste em blocos*. 1999. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística, USP, São Paulo, 1999.
100. 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, p. 605-610, 1987.
101. SIMONOFF, J. S. *Smoothing methods in statistics*. New York : Springer, 1996.
102. SINGER, J.M. ; ANDRADE, D.F. Regression models for the analysis of Pre-test/Posttest data. *Biometrics*, v. 53, p. 729-735, 1997.
103. SINGER, J. M. ; ANDRADE, D. F. Analysis of longitudinal data. In: SEN, P. K.; RAO, C. R. (Ed.) *Bioenvironmental and Public Health Statistics*. Amsterdam : Elsevier, 2000. p. 115-160. (Handbook of Statistics, 18)
104. SINGER, J. M ; NOBRE, J. S ; ROCHA, F. M. M. *Análise de dados longitudinais*. São Paulo : IME-USP, 2012. (Versão parcial preliminar)
105. SORENSEN, D. A. *Gibbs sampling in quantitative genetics*. Koppenhagen: Danish Institute of Animal Science, 1996.
106. SPIEGELHALTER, D. J. et al. A Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, v.64, p. 583-639, 2002.

107. STRAM, D.O. ; LEE, J.W. Variance components testing in the longitudinal mixed effects model. *Biometrics*, v. 50, p.1171-1177, 1994.
108. SUYAMA, E. *Modelos de efeitos aleatórios para dados longitudinais*. In : 40^a RBRAS e 6^a SEAGRO, 1995. Ribeirão Preto, SP, 1995. 94 p.
109. TAN, F. E. S.; OUWENS, M. J. N. ; BERGER, M. P. F. Detection of influential observations in longitudinal mixed effects regression models. *The Statistician*, v. 50, p. 271-284, 2001.
110. TOUNTENBURG, H. *Prior information in linear models*. New York : Wiley, 1982.
111. VENEZUELA, M. K. *Modelos lineares generalizados para análise de dados com medidas repetidas*. 2003. Dissertação (Mestrado em Estatística) Instituto de Matemática e Estatística, USP, São Paulo, 2003.
112. VENEZUELA, M. K.; SANDOVAL, M. C.; BOTTER, D. A. Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation*, v. 77, p. 879-888, 2007.
113. VENEZUELA, M. K.; SANDOVAL, M. C.; BOTTER, D. A. Local influence in estimating equations. *Computational Statistics Data Analysis*, v. 55, p.1867-1883, 2010.
114. VERBEKE, G. ; LESAFFRE, E. *Large Samples properties of the maximum likelihood estimators in linear mixed models with misspecified random effects distributions*. Belgium : Catholic University of Leuven, Biostatistical Centre for Clinical Trials, 1996. (Technical report)
115. VERBEKE, G. ; MOLENBERGHS, G. *Linear mixed models in practice : a SAS-oriented approach*. New York : Springer-Verlag, 1997.
116. VERBEKE, G. ; MOLENBERGHS, G. *Models for discrete longitudinal data*. New York : Springer-Verlag, 2000.
117. VERBEKE, G. ; MOLENBERGHS, G. The Use of score tests for inference on variance components. *Biometrics*, v. 59, p. 254-262, 2003.
118. VERBEKE, G. ; MOLENBERGHS, G. *Models for discrete longitudinal data*. New York : Springer Science, 2005.

119. VONESH, E. F. ; CHINCHILLI, V. M. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, v. 52, n. 2, p.572-587, 1996.
120. VONESH, E. F. ; CHINCHILLI, V. M. *Linear and non-linear models for the analysis of repeated measurements*. New York : Marcel Dekker, 1997.
121. WAHBA, G. *Spline models for observational data*. Philadelphia: SIAM, 1990.
122. WEI, B.C.; HU, Y.Q. ; FUNG, W.K. Generalized leverage and its applications. *Scandinavian Journal of Statistics*, v. 25, p. 25-37, 1998.
123. WU, H. ; ZHANG, J. T. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. New Jersey : Wiley-Interscience, 2006.