Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos lineares mistos em dados longitudinais com o uso do
pacote ASReml-R

Renata Alcarde

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2012

Renata Alcarde
Licenciada em Matemática

**Modelos lineares mistos em dados longitudinais com o uso do pacote
ASReml-R**

Orientadora:
Prof^a Dr^a **SÔNIA MARIA DE STEFANO PIEDADE**

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2012**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Alcarde, Renata

Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R /
Renata Alcarde. - - Piracicaba, 2012.
156 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Análise de dados longitudinais 2. Aviários 3. Estatística aplicada 4. Eucalipto
5. Experimentos agrícolas 6. Modelos lineares mistos 7. Verossimilhança I. Título

CDD 519.5
A348m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

A minha família,

meus pais, Antonio Sergio e Olga,
meu irmão Rodrigo ,
minha tia Maria Inês,
e meu namorado Fellipe.

Dedico.

AGRADECIMENTOS

A Deus pela presença constante em minha vida, por ajudar a trilhar meus caminhos. A Escola Superior de Agricultura “Luiz de Queiroz” e ao Programa de Pós-Graduação em Estatística e Experimentação Agronômica, pela oportunidade e conhecimentos adquiridos. A professora Sônia Maria De Stefano Piedade, por me orientar não apenas neste trabalho, mas na vida. Pela confiança, pelos conselhos, por todos os sorrisos e abraços.

A professora Clarice Garcia Borges Demétrio, pelos ensinamentos e pela disposição em me ajudar. Ao professor Chris Brien pelas sugestões. Ao professor César Gonçalves de Lima, pela companhia e ao professor Amauri de Almeida Machado, por me ensinar o poder das palavras “Por quê?”.

Aos professores do Departamento de Ciências Exatas: Décio Barbin, Carlos Tadeu dos Santos Dias, Edwin Moises Marcos Ortega, Gabriel Adrian Sarries, Roseli Aparecida Leandro, Silvio Sandoval Zocchi, Taciana Villela Savian, pelo conhecimento compartilhado.

As secretárias Luciane Brajão e Solange de Assis Paes Sabadin, por estarem sempre prontas a ajudar sem ver a quem.

Aos funcionários, Eduardo Bonilha, Jorge Alexandre Wiendl e Rosni Honofre Aparecido Pinto, pelo suporte, e a Luciene, por manter nossos ambientes agradáveis e propícios ao estudo.

Aos funcionários do departamento de Economia, Administração e Sociologia, Pedro Scardua e Maria Aparecida Zani Ribeiro, pela gentileza com que sempre me atenderam.

Às revisoras, Eliana Maria Garcia, Maria da Glória Eloi da Silva, Rosa e Samira Kraide, pela atenção e presteza.

À Marina Rodrigues Maestre e à Simone Daniela Sartorio, pela amizade, pelo companheirismo, pela presença nos momentos felizes e não tão felizes e pelos abraços. Ao Edilan Sant’Ana Quaresma, pelas palavras amigas. À Lucimery Afonso dos Santos, Vanderly Janeiro e Wilson Alves de Oliveira, amigos queridos que sempre levarei em meu coração.

À Angela Mello Coelho, Fernanda Bühner Rizzato, Mirian Fernandes Carvalho Araujo, Lúcio Borges de Araujo, por serem exemplos de perseverança e conquista. À Bianca, pela oportunidade de conhecê-la.

A todos os amigos da sala de estudos, pelos momentos de descontração e pelos

cafezinhos.

Ao CNPq pela concessão da bolsa de estudos.

Agradecimento especial

À minha família. Meus pais, Antonio Sergio e Olga, meu irmão Rodrigo e minha tia Maria Inês, pelo incentivo, por estarem sempre ao meu lado, por se alegrarem e chorarem comigo, por serem referências em minha vida.

Ao Felliipe, meu namorado, por me apoiar em minhas decisões, me compreender e pela constante presença.

A todos vocês, meu queridos, obrigada por fazerem parte da minha vida!

Caminheiro, você sabe, não existe caminho.

*Passo a passo, pouco a pouco,
e o caminho se faz.*

Benedito B. Prado

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	19
1 INTRODUÇÃO	21
2 REVISÃO BIBLIOGRÁFICA	25
2.1 Definições e notações	25
2.2 Diagrama de aleatorização	27
2.3 Diagramas de Hasse	29
2.4 Modelos Mistos	33
2.5 Estruturas de covariância	35
2.6 Estimação dos componentes de variância	38
2.6.1 Método ANOVA	39
2.6.2 Máxima Verossimilhança	39
2.6.3 Máxima Verossimilhança Restrita	40
2.7 Estimação e predição	43
2.7.1 Estimação dos termos fixos	43
2.7.2 Predição dos termos aleatórios	44
2.7.3 Estimação dos termos fixos e predição dos termos aleatórios simultaneamente	44
2.8 Suavização spline cúbica	46
2.8.1 Suavização Spline cúbica: um único indivíduo	46
2.8.2 Suavização Spline cúbica: vários indivíduos	47
2.8.3 Suavização Spline cúbica: presença de fatores qualitativos	48
2.8.4 Representação da análise da variância	49
2.8.5 Suavização Spline cúbica: dados longitudinais	50
2.9 Construção de modelos mistos	51
2.10 Seleção de modelos	56
2.10.1 Teste da razão de verossimilhanças	57
2.10.1.1 Inferência para parâmetros de efeitos fixos	57
2.10.1.2 Inferência para os parâmetros de efeitos aleatórios	58
2.10.2 Testes alternativos para parâmetros de efeito fixo	59
2.10.2.1 Teste de Wald	59

2.10.2.2	Teste t	59
2.10.2.3	Teste Wald- F	60
2.10.3	Cr�terios de informa��o	60
2.10.4	Estrat�gias de sele��o de modelos	61
2.10.4.1	Estrat�gia <i>top-down</i>	61
2.10.4.2	Estrat�gia <i>step-up</i>	61
2.11	Diagn�sticos	62
2.11.1	Res�duos	62
2.11.2	Distribui��o dos res�duos e dos efeitos aleat�rios	63
2.11.3	Res�duos com confundimento m�nimo	64
3	MATERIAL E M�TODOS	69
3.1	Material	69
3.1.1	Experimento com instala��es av�colas	69
3.1.2	Experimento com polpas de Eucalipto	69
3.2	M�todos	70
3.3	EBLUP com confundimento m�nimo	72
3.4	Pacote estat�stico	74
3.4.1	ASReml	74
3.4.2	Novas fun��es	75
4	RESULTADOS E DISCUSS�O	79
4.1	Experimento com instala��es av�colas	79
4.2	Experimento com polpas de Eucalipto	92
5	CONSIDERA��ES FINAIS	115
	REFER�NCIAS	117
	ANEXOS	121

RESUMO

Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R

Grande parte dos experimentos instalados atualmente é planejada para que sejam realizadas observações ao longo do tempo, ou em diferentes profundidades, enfim, tais experimentos geralmente contêm um fator longitudinal. Uma maneira de se analisar esse tipo de conjunto de dados é utilizando modelos mistos, por meio da inclusão de fatores de efeito aleatório e, fazendo uso do método da máxima verossimilhança restrita (REML), podem ser estimados os componentes de variância associados a tais fatores com um menor viés. O pacote estatístico ASReml-R, muito eficiente no ajuste de modelos lineares mistos por possuir uma grande variedade de estruturas para as matrizes de variâncias e covariâncias já implementadas, apresenta o inconveniente de não ter como objetos as matrizes de delineamento $\mathbf{X}$ e $\mathbf{Z}$, nem as matrizes de variâncias e covariâncias $\mathbf{D}$ e $\mathbf{\Sigma}$, sendo estas de grande importância para a verificação das pressuposições do modelo. Este trabalho reuniu ferramentas que facilitam e fornecem passos para a construção de modelos baseados na aleatorização, tais como o diagrama de Hasse, o diagrama de aleatorização e a construção de modelos mistos incluindo fatores longitudinais. Sendo o vetor de resíduos condicionais e o vetor de parâmetros de efeitos aleatórios confundidos, ou seja, não independentes, foram obtidos resíduos, denominados na literatura, resíduos com confundimento mínimo e, como proposta deste trabalho foi calculado o EBLUP com confundimento mínimo. Para tanto, foram implementadas funções que, utilizando os objetos de um modelo ajustado com o uso do pacote estatístico ASReml-R, tornam disponíveis as matrizes de interesse e calculam os resíduos com confundimento mínimo e o EBLUP com confundimento mínimo. Para elucidar as técnicas neste apresentadas e salientar a importância da verificação das pressuposições do modelo adotado, foram considerados dois exemplos contendo fatores longitudinais, sendo o primeiro um experimento simples, visando a comparação da eficiência de diferentes coberturas em instalações avícolas, e o segundo um experimento realizado em três fases, contendo fatores inteiramente confundidos, com o objetivos de avaliar características do papel produzido por diferentes espécies de eucaliptos em diferentes idades.

Palavras-chave: Diagrama de Hasse; Diagrama de aleatorização; Formulação de modelos mistos; Resíduos com confundimento mínimo; EBLUP com confundimento mínimo

ABSTRACT

Linear Mixed Models with longitudinal data using ASReml-R package

Currently, most part of the experiments installed is designed to be carried out observations over time or at different depths. These experiments usually have a longitudinal factor. One way of analyzing this data set is by using mixed models through means of inclusion of random effect factors, and it is possible to estimate the variance components associated to such factors with lower bias by using the Restricted maximum likelihood method (REML). The ASReml-R statistic package, very efficient in fitting mixed linear models because it has a wide variety of structures for the variance - covariance matrices already implemented, presents the disadvantage of having neither the design matrices $\mathbf{X}$ and $\mathbf{Z}$, nor the variance - covariance matrices $\mathbf{D}$ and $\mathbf{\Sigma}$, and they are very important to verify the assumption of the model. This paper gathered tools which facilitate and provide steps to build models based on randomization such as the Hasse diagram, randomization diagram and the mixed model formulations including longitudinal factors. Since the conditional residuals and random effect parameters are confounded, that is, not independent, it was calculated residues called in the literature as least confounded residuals and as a proposal of this work, it was calculated the least confound EBLUP. It was implemented functions which using the objects of fitted models with the use of the ASReml-R statistic package becoming available the matrices of interests and calculate the least confounded residuals and the least confounded EBLUP. To elucidate the techniques shown in this paper and highlight the importance of the verification of the adopted model's assumptions, it was considered two examples with longitudinal factors. The former example was a simple experiment and the second one conducted in three phases, containing completely confounded factors, with the purpose of evaluating the characteristics of the paper produced by different species of eucalyptus from different ages.

Keywords: Hasse diagram; Randomization diagram; Mixed model formulations; Least confounded residuals; Least confounded EBLUP

LISTA DE FIGURAS

Figura 1 - Exemplo de um diagrama de aleatorização para um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis	28
Figura 2 - Diagramas de Hasse para obtenção dos números de graus de liberdade, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis	30
Figura 3 - Diagramas de Hasse para obtenção das matrizes núcleo das formas quadráticas, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis	31
Figura 4 - Diagramas de Hasse para obtenção das esperanças dos quadrados médios, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis	32
Figura 5 - Diagrama de aleatorização para o experimento com instalações avícolas	79
Figura 6 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com instalações avícolas, considerando-se o modelo IR	80
Figura 7 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com instalações avícolas, considerando-se o modelo F	80
Figura 8 - Gráfico de perfis individuais e médios para o experimento com instalações avícolas	83
Figura 9 - Gráfico dos valores para o EBLUP (modelo 40) para o experimento com instalações avícolas	87
Figura 10 - Gráfico dos resíduos condicionais estudantizados (modelo 40) para o experimento com instalações avícolas	87
Figura 11 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 40) para o experimento com instalações avícolas	88
Figura 12 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 40) para o experimento com instalações avícolas	88
Figura 13 - Gráfico dos valores para o EBLUP (modelo 41) para o experimento com instalações avícolas	89

Figura 14 - Gráfico dos resíduos condicionais estudentizados (modelo 41) para o experimento com instalações avícolas	90
Figura 15 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 41) para o experimento com instalações avícolas	90
Figura 16 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 41) para o experimento com instalações avícolas	91
Figura 17 - Curvas ajustadas de acordo com o modelo 41 para o experimento com instalações avícolas	92
Figura 18 - Diagrama de aleatorização para o experimento com polpas de Eucalipto .	93
Figura 19 - Diagramas de Hasse para obtenção dos números de graus de liberdade para o experimento com polpas de Eucalipto, considerando-se a primeira fase (camada das amostras) e a segunda fase (camada das posições) do experimento	94
Figura 20 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com polpas de Eucalipto, considerando-se a terceira fase do experimento (camadas dos grupos e dos tempos)	95
Figura 21 - Gráfico de perfis individuais e médios para o experimento com polpas de Eucalipto	97
Figura 22 - Gráfico dos valores para o EBLUP (modelo 51) para o experimento com polpas de Eucalipto	103
Figura 23 - Gráfico dos resíduos condicionais estudentizados (modelo 51) para o experimento com polpas de Eucalipto	103
Figura 24 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 51) para o experimento com polpas de Eucalipto	104
Figura 25 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 51) para o experimento com polpas de Eucalipto	104
Figura 26 - Gráfico dos valores para o EBLUP (modelo 52) para o experimento com polpas de Eucalipto	106
Figura 27 - Gráfico dos resíduos condicionais estudentizados (modelo 52) para o experimento com polpas de Eucalipto	106
Figura 28 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 52) para o experimento com polpas de Eucalipto	107

Figura 29 -Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 52) para o experimento com polpas de Eucalipto	107
Figura 30 -Curvas ajustadas de acordo com o modelo 52 para o experimento com polpas de Eucalipto	108
Figura 31 -Gráfico dos valores para o EBLUP (modelo 53) para o experimento com polpas de Eucalipto	109
Figura 32 -Gráfico dos resíduos condicionais estudentizados (modelo 53) para o expe- rimento com polpas de Eucalipto	110
Figura 33 -Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 53) para o experimento com polpas de Eucalipto	110
Figura 34 -Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 53) para o experimento com polpas de Eucalipto	111
Figura 35 -Curvas ajustadas de acordo com o modelo 53 para o experimento com polpas de Eucalipto	112

LISTA DE TABELAS

Tabela 1 - Notação resumo para modelos mistos	27
Tabela 2 - Estruturas, fórmulas estruturais e fontes de variação para um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis	29
Tabela 3 - Tabela da análise da variância contendo os números de graus de liberdade, as expressões para as somas de quadrados e para as esperanças dos quadrados médios de cada fonte de variação	33
Tabela 4 - Decomposição da tabela da análise da variância, considerando um experimento fatorial, em que o fator A é qualitativo e o fator T quantitativo	50
Tabela 5 - Resumo para os operadores de tendência e de correlação desigual	56
Tabela 6 - Decomposição dos números de graus de liberdade para o experimento com instalações avícolas	81
Tabela 7 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com instalações avícolas	85
Tabela 8 - Testes de Wald- F para os efeitos fixos presentes no modelo 39, do experimento com instalações avícolas	86
Tabela 9 - Decomposição dos números de graus de liberdade para o experimento com polpas de Eucalipto	95
Tabela 10 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com polpa de Eucalipto	100
Tabela 11 - Testes de Wald- F para os efeitos fixos presentes no modelo 49, do experimento com polpas de Eucalipto	101
Tabela 12 - Testes de Wald F para os efeitos fixos presentes no modelo 50, do experimento com polpas de Eucalipto	102
Tabela 13 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com polpa de Eucalipto	102
Tabela 14 - Testes de Wald- F para os efeitos fixos presentes no modelo 52, do experimento com polpas de Eucalipto	105
Tabela 15 - Critérios de informação para comparação dos modelos 52 e 53	112
Tabela 16 - Dados do experimento com polpas de Eucalipto	132

1 INTRODUÇÃO

Medidas repetidas é o termo adotado para descrever experimentos nos quais uma ou mais variáveis resposta são observadas em cada unidade experimental em mais de uma ocasião, e em alguns casos, sob diferentes condições experimentais. Tais experimentos são amplamente utilizados em áreas agrícolas, biológicas, médicas, dentre outras.

Um caso particular de medidas repetidas é denominado dados longitudinais, no qual o comportamento da variável resposta é observado ao longo de uma direção específica, sendo o tempo a mais comum. Uma característica inerente aos dados longitudinais é a ausência da independência entre as observações, uma vez que as mesmas são realizadas em uma mesma unidade experimental, outra possível característica é a heterogeneidade de variâncias nas diversas ocasiões observadas.

A abordagem que utiliza os modelos mistos em dois estágios, proposta por Laird e Ware (1982), fazendo uso das idéias de Harville (1977), revolucionou a análise desse tipo de dados, pois a família de modelos proposta possibilitou variações nas estruturas das matrizes de variâncias e covariâncias das observações entre unidades e dentro das unidades experimentais.

Experimentos mais complexos, como pode ser o caso dos experimentos envolvendo fatores longitudinais, podem possuir efeitos de alguns fatores confundidos, sendo difícil a estimação de parâmetros associados aos mesmos. Para que sejam identificados tais confundimentos aconselha-se a utilização de duas técnicas: o diagrama de Hasse e o diagrama de aleatorização.

O diagrama de Hasse, embora desenvolvido para experimentos ortogonais, é uma ferramenta que auxilia na visualização da estrutura do experimento, facilitando assim a compreensão dos confundimentos e a obtenção dos números de graus de liberdade associados a cada fonte de variação. A segunda técnica mencionada, o diagrama de aleatorização, permite a visualização das camadas presentes em cada experimento e dos fatores associados as mesmas. Pela própria construção do diagrama de aleatorização pode-se saber quais fatores foram aleatorizados (ou atribuídos de modo sistemático) a quais outros fatores, tornando claro o confundimento, caso exista.

Muitas vezes, a formulação de modelos para experimentos contendo dados longitudinais não é simples, devido ao número grande de fatores e de seus respectivos níveis, o que justifica o desenvolvimento de uma técnica para encontrar um modelo inicial sob pesquisa. A técnica proposta por Brien e Demétrio (2009) fornece passos para a construção

do modelo baseado na aleatorização, sendo assim um grande facilitador de tal processo.

Tendo em vista um modelo inicial a ser ajustado, faz-se necessária a escolha de um método para a estimação dos componentes de variância, uma vez que os modelos mistos incorporam termos aleatórios adicionais ao modelo. Neste trabalho o método utilizado foi o da máxima verossimilhança restrita, que fornece estimativas com menor viés, juntamente com o algoritmo iterativo Average Information, que reduz o tempo de processamento. E, para estimar o vetor de parâmetros de efeitos fixos, foi utilizado o método dos mínimos quadrados generalizados.

O modelo misto, tal como utilizado neste trabalho, tem como pressuposições a normalidade dos resíduos e a normalidade do vetor de parâmetros de efeito aleatório. Ajustado um modelo, a verificação da normalidade dos resíduos foi feita utilizando-se os resíduos com confundimento mínimo, propostos por Hilden-Minton (1995), que consiste de uma transformação linear do vetor de resíduos condicionais. Seguindo a proposta dos resíduos com confundimento mínimo, foi desenvolvido neste trabalho o EBLUP com confundimento mínimo, por meio dos quais verificou-se a normalidade do vetor de parâmetros de efeito aleatório fazendo uso do gráfico quantil-quantil.

Cabe salientar que todos os modelos foram ajustados utilizando o pacote estatístico ASReml-R, para o software R. Este pacote, apesar de existir um custo para se obter sua licença, contém uma gama de opções para as estruturas das matrizes de variâncias e covariâncias, bem como os modelos de suavização spline cúbica, já programados.

Para exemplificar o uso das técnicas propostas foram analisados dois conjuntos de dados. O primeiro oriundo de uma pesquisa realizada por Castro (2011), com o objetivo de avaliar o conforto térmico e o bem estar de animais em instalações avícolas. Neste experimento foram comparados dois tipos de telha associados a dois tipos de cobertura. Várias características foram avaliadas, porém a variável resposta analisada neste trabalho foi temperatura de bulbo seco ($^{\circ}\text{C}$), com medições realizadas em quatro horas distintas ao longo do dia.

O segundo conjunto de dados analisado é fruto da pesquisa de Pereira (1969), o qual se trata de um experimento mais complexo, realizado em três fases, com efeitos de fatores confundidos. O objetivo do experimento foi avaliar duas diferentes espécies de Eucalipto em duas idades quanto as características do papel produzido, considerando-se seis tempos de refinação das polpas produzidas. A característica avaliada como exemplo foi peso específico aparente (g/cm^3).

Este trabalho teve como objetivos, além do desenvolvimento do EBLUP com confundimento mínimo, reunir as ferramentas: diagrama de Hasse, diagrama de aleatorização, a construção de modelos mistos envolvendo dados longitudinais, resíduos com confundimento mínimo e EBLUP com confundimento mínimo, de tal modo a fornecer passos para a análise de dados com fatores longitudinais. Outro objetivo deste trabalho foi a implementação de funções, no software estatístico R, para obtenção das matrizes $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{D}$ e $\mathbf{\Sigma}$, bem como do vetor de resíduos com confundimento mínimo e EBLUP com confundimento mínimo, fazendo uso do pacote estatístico ASReml-R.

Na seção que se segue é feita uma revisão da literatura a respeito do assunto abordado, na Seção 3 são descritos os dois conjuntos de dados utilizados para demonstração da técnica proposta. Nesta seção é também apresentado o EBLUP com confundimento mínimo e as funções construídas para a obtenção das matrizes $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{D}$ e $\mathbf{\Sigma}$, do vetor de resíduos com confundimento mínimo e do vetor EBLUP com confundimento mínimo, juntamente com seus respectivos argumentos. Na Seção 4 são realizadas as análises dos dois experimentos e na Seção 5 são apresentadas as considerações finais sobre o trabalho.

2 REVISÃO BIBLIOGRÁFICA

Nesta seção serão apresentadas algumas definições, notações e os principais conceitos relacionados aos modelos mistos e dados longitudinais. Como se obter um modelo misto, bem como a estimação dos parâmetros de efeito fixo e a predição dos de efeito aleatório. Seguindo a proposta dos modelos mistos, serão abordados os modelos de suavização spline cúbica, como uma alternativa aos modelos não lineares.

Será apresentada também uma proposta de formulação de modelos mistos, incluindo dados longitudinais encontrada na literatura. E, para a verificação da normalidade do vetor de resíduos serão construídos os resíduos com confundimento mínimo.

2.1 Definições e notações

Bailey (2005), Machado et al. (2005), Brien (2007) entre outros autores, apresentam algumas definições importantes para o desenvolvimento, compreensão e análise de experimentos.

Para um melhor entendimento de tais definições, considere um experimento casualizado em blocos, com b o número de blocos, e esquema de tratamento fatorial, em que o fator qualitativo A , possui a níveis, e o fator qualitativo C , possui c níveis.

Definição 1. Fatores não aleatorizados são aqueles intrínsecos ao experimento. De modo geral fatores de controle local são fatores não aleatorizados. **Fatores aleatorizados** são aqueles associados às unidades de observação como um resultado da aleatorização. Fatores de tratamento podem ser vistos como fatores aleatorizados. **Fatores longitudinais** são aqueles para os quais nenhum fator é aleatorizado e que indexam as sucessivas observações em uma mesma unidade experimental, como é o caso do tempo, ou profundidade. *No exemplo:* Blocos e Parcelas são fatores não aleatorizados, os fatores A e C são aleatorizados. O exemplo não possui fatores longitudinais.

Definição 2. Dois fatores são considerados cruzados se cada nível de um fator, combinado com os níveis de outro fator, é sempre o mesmo, ou seja, conserva uma característica associada ao nível do fator.

No exemplo: os fatores A e C são cruzados.

Notação: $A * C \Rightarrow A + C + A\#C$, com o cruzamento dos fatores A e C tem-se os efeitos principais dos fatores A e C , e o efeito da interação $A\#C$.

Definição 3. Dois fatores são considerados aninhados se cada nível de um fator, combinado com os níveis de outro, é sempre distinto, ou seja, se o mesmo nível de um fator aninhado, a partir de diferentes níveis do fator que o aninha, não possuem características em comum.

No exemplo: o fator Parcelas está aninhado ao fator Blocos.

Notação: Blocos/Parcelas $\Rightarrow$ Blocos + Parcelas[Blocos], com essa relação tem-se os efeitos de Blocos e de Parcelas dentro de Blocos, denotado por Parcelas[Blocos].

Definição 4. Um fator generalizado é um fator formado a partir de um ou mais fatores (originais) e cujos níveis são combinações que ocorrem no experimento, trata-se de uma lista de fatores, separados pelo operador conjunção lógica (*wedge* - $\wedge$).

No exemplo: São fatores generalizados Blocos, Blocos $\wedge$ Parcelas, A, C e A $\wedge$ C, respectivos aos efeitos de Blocos, Parcelas[Blocos], A, C e A $\#$ C.

Definição 5. Um fator generalizado, F_1 , é dito marginal em relação a outro fator generalizado, F_2 , se todos os fatores simples que compõem F_1 formam um subconjunto dos fatores simples que compõem F_2 .

No exemplo: o fator generalizado Blocos é marginal ao fator generalizado Blocos $\wedge$ Parcelas. Os fatores generalizados A e C são marginais ao fator generalizado A $\wedge$ C.

O fator universal, U, que representa a média geral, é marginal a todo fator generalizado presente na estrutura.

Na Tabela 1 encontram-se as notações para os principais operadores envolvendo dois ou mais fatores, suas relações e as expansões de suas fórmulas estruturais, que são representações algébricas das relações entre os fatores e fornecem um conveniente modo de tornar o modelo mais compacto. Nessa tabela, A, B e C são fatores e L, M e N são fórmulas para modelos.

Tabela 1 - Notação resumo para modelos mistos

1. Operador conjunção lógica, *wedge* ($\wedge$)

- (i) $A \wedge B = B \wedge A$ é o fator generalizado formado pelas combinações dos níveis de A e B
- (ii) $(A \wedge B) \wedge C = A \wedge (B \wedge C)$
- (iii) $(A \wedge C) \wedge (B \wedge C) = A \wedge B \wedge C$
- (iv) $L \wedge M$ é expandido, tomando todos os produtos de pares dos elementos expandidos de L e M

2. Operador aditivo (+)

- (i) $(A+B) \wedge (C+D) = A \wedge C + A \wedge D + B \wedge C + B \wedge D$

3. Operador fator generalizado [gf($\cdot$)]

- (i) Forma o fator generalizado de todos os fatores presentes no modelo que é seu argumento
- (ii) $gf(L) = A \wedge B \wedge C \wedge \dots$, em que A, B, C, ... são fatores em L
- (iii) $(A + B)/C = A + B + gf(A + B) \wedge C = A + B + A \wedge B \wedge C$

4. Operador de aninhamento (/)

- (i) $L/M = L + gf(L) \wedge M$
- (ii) $L/(M/N) = (L/M)/N$
- (iii) $L/(M + N) = (L/M) + (L/N)$

5. Operador de cruzamento (*)

- (i) $L * M = L + M + L \wedge M$
- (ii) $L * (M + N) = (L * M) + (L * N)$
- (iii) $L * M/N = L * (M/N)$

Fonte: Brien e Demétrio (2009)

2.2 Diagrama de aleatorização

Num modelo misto a distinção entre fatores de controle local e fatores de tratamento, denominados por Brien (1983) e Brien e Bailey (2006), dentre tantos outros, como fatores não aleatorizados e aleatorizados, respectivamente, facilitam a construção da tabela da análise da variância, exibindo todos os confundimentos, e a inclusão de termos

devido à aleatorização.

Brien (1983) apresenta o conceito de camada (*tier*), como sendo um agrupamento de fatores, indexando um dos conjuntos de objetos envolvidos em uma aleatorização. Fatores em diferentes camadas são associados por uma aleatorização. Experimentos *two-tiered* são aqueles que apresentam duas camadas e apenas um nível de aleatorização, geralmente, em tais experimentos, uma camada se refere aos fatores não aleatorizados e outra aos fatores aleatorizados. Experimentos que envolvem múltiplas aleatorizações e possuem mais que duas camadas são denominados *multitiered*.

Brien e Bailey (2006) desenvolveram um sistema para descrever as aleatorizações em termos dos conjuntos de objetos, suas camadas associadas e os fatores aninhados, usando diagramas de aleatorização, os quais resumem a aleatorização em um experimento e facilitam a compreensão do confundimento entre os fatores.

Nos diagramas de aleatorização, cada painel exibe uma lista de fatores referentes a cada camada, seus números de níveis e as relações de aninhamento, sendo escritos com a letra inicial maiúscula. Fora de cada painel são escritos o número e o nome dos objetos, em letra minúscula. Uma seta, da esquerda para a direita, indica que o fator à esquerda está sendo aleatorizado ao fator da direita. Caso não haja aleatorização e sim uma atribuição sistemática, a seta deve ser tracejada.

Quando se deseja indicar a combinação dos níveis de dois ou mais fatores em uma mesma camada, utiliza-se o símbolo ● com duas ou mais linhas. Para os casos em que os fatores a terem seus níveis combinados estão em camadas distintas, o símbolo utilizado é ■. Os autores definem símbolos adicionais para os casos em que apenas uma fração das combinações dos níveis dos fatores é aleatorizada, ou atribuída de modo sistemático, para delineamentos ortogonais, dentre outros.

Considerando o experimento descrito como exemplo, o diagrama de aleatorização é apresentado na figura a seguir.

Figura 1 - Exemplo de um diagrama de aleatorização para um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis

2.3 Diagramas de Hasse

O diagrama de Hasse é um conjunto parcialmente ordenado, em que a ordem parcial é estabelecida pela relação de aninhamento existente entre os fatores generalizados.

Taylor e Hilton (1981) afirmam que o diagrama fornece uma valiosa perspectiva complementar para a análise de variância e as técnicas de análise, por meio de uma conexão entre a descrição verbal do experimento e o correspondente modelo linear estatístico. Trata-se de uma poderosa ferramenta visual utilizada na representação da estrutura dos fatores de um delineamento experimental ortogonal (MACHADO et al., 2005).

Brien (2007) desenvolveu algumas regras para construção do diagrama de Hasse, bem como para a obtenção dos respectivos números de graus de liberdade, das matrizes núcleo das formas quadráticas (Q) referentes às somas de quadrados ($y'Qy$), e das esperanças dos quadrados médios, apresentadas a seguir:

Regra 1: Determinar se um fator é aleatorizado ou não aleatorizado.

No exemplo:

Fatores não aleatorizados: Blocos e Parcelas;

Fatores aleatorizados: A e C

Regra 2: Determinar a estrutura experimental, descrevendo as relações de aninhamento ou de cruzamento entre os fatores não aleatorizados. Em seguida, descrever as relações de aninhamento ou cruzamento entre os fatores aleatorizados e entre os fatores aleatorizados e não aleatorizados, simultaneamente.

No exemplo:

Tabela 2 - Estruturas, fórmulas estruturais e fontes de variação para um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis

Estrutura	Fórmula	Fontes de variação
não aleatorizados	b Blocos/ ac Parcelas	Blocos + Parcelas[Blocos]
aleatorizados	a A * c C	A + C + A#C

Regra 3: O diagrama de Hasse para os fatores generalizados de uma fórmula estrutural é formado tal que, a posição dos pontos representando tais fatores generalizados indique a relação entre os fatores presentes em cada fórmula estrutural. Desse modo, a representação de um fator generalizado deve estar posicionada acima da representação para o fator

generalizado para o qual este é marginal. Se dois fatores são cruzados então os fatores generalizados para os efeitos principais devem estar no mesmo nível no diagrama. Acima de todos os fatores, coloca-se o fator universal, U , representando a média geral. À esquerda do ponto escreve-se o fator generalizado e o número de níveis e à direita, escreve-se o número de graus de liberdade e o respectivo efeito. O número de graus de liberdade é obtido pela diferença entre o número de níveis e a soma dos números de graus de liberdade de todos os fatores marginais ao fator em questão.

No exemplo:

Figura 2 - Diagramas de Hasse para obtenção dos números de graus de liberdade, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis

Regra 4: A expressão para a matriz Q para cada fator generalizado, em termos das matrizes M , em que M é a matriz de projeção dada por $X(X'X)^{-1}X'$ e X é a matriz de incidência de posto completo, é obtida substituindo-se, no diagrama de Hasse, o número de níveis de cada fator generalizado pela respectiva matriz Q e do lado direito no diagrama, as expressões são obtidas pela diferença entre a matriz M em questão e a soma das expressões das matrizes Q dos fatores marginais a este fator generalizado.

No exemplo:

Figura 3- Diagramas de Hasse para obtenção das matrizes núcleo das formas quadráticas, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis

Regra 5: Para se obterem os modelos de esperança e variância, deve-se determinar se um fator generalizado é, potencialmente, de esperança ou variação. Se um fator generalizado envolve somente fatores de efeito fixo, então este é um termo potencial de esperança e, caso contrário será um termo de variação. O termo consistindo de todos os fatores não aleatorizados é designado como aleatório. O modelo de esperança maximal (ψ) dá-se pela soma de todos os termos potenciais de esperança, exceto aqueles marginais a outro termo de esperança. Caso não haja termo de esperança, tal modelo terá um único termo, representando a média geral. E, o modelo de variação maximal dá-se pela soma de todos os termos de variação.

No exemplo: Supondo os fatores A e C de efeitos fixos e o fator Blocos de efeito aleatório, tem-se que o modelo de esperança maximal é dado, simbolicamente, por

$$\psi = E(Y) = A \wedge C,$$

e o modelo de variação é dado, simbolicamente, por:

$$\text{Var}(Y) = \text{Blocos} + \text{Blocos} \wedge \text{Parcelas}.$$

Regra 6: Os passos para a construção das esperanças dos quadrados médios, para um experimento ortogonal, são:

1. Para cada fórmula estrutural, tomar o diagrama de Hasse para os fatores generalizados e, para cada fator generalizado F , substituir o número de níveis, f , por $(n/f)\sigma_F^2 - n$ é o número de unidades experimentais - se F é um termo no modelo de variação, ou

por $q_F(\Psi)$ se F é um termo no modelo de esperança. Do lado direito de cada fator generalizado, informar sua contribuição para a esperança do quadrado médio, incluindo a expressão à esquerda de F e à esquerda de todo fator generalizado para o qual F é marginal. O procedimento é realizado de baixo para cima.

2. Adicionar as contribuições dos fatores não aleatorizados, calculadas no diagrama de Hasse, às esperanças dos quadrados médios, na tabela de análise da variância, colocando cada contribuição referente a sua fonte na tabela, a menos que a fonte tenha sido subdividida, nesse caso colocar tal contribuição em cada partição.
3. Repetir o passo 2. para a outra fórmula estrutural, adicionando as contribuições àquelas que já estão na tabela. Caso o fator ocorra mais de uma vez no diagrama de Hasse, sua contribuição será adicionada apenas uma vez na tabela.

No exemplo: os diagramas de Hasse para a obtenção das esperanças dos quadrados médios de cada fonte de variação, são dados por:

Figura 4 - Diagramas de Hasse para obtenção das esperanças dos quadrados médios, considerando-se um experimento casualizado em blocos, com b blocos, e esquema de tratamento fatorial, em que o fator A possui a níveis e o fator C , c níveis

E a tabela da análise da variância é apresentada conforme segue:

Tabela 3 - Tabela da análise da variância contendo os números de graus de liberdade, as expressões para as somas de quadrados e para as esperanças dos quadrados médios de cada fonte de variação

Fontes de Variação	gl	SQ	E(QM)
Blocos	$b - 1$	$\mathbf{y}'\mathbf{Q}_B\mathbf{y}$	$\sigma_{BP}^2 + ac \sigma_B^2$
Parcelas[Blocos]	$b(ac - 1)$	$\mathbf{y}'\mathbf{Q}_{BP}\mathbf{y}$	
A	$a - 1$	$\mathbf{y}'\mathbf{Q}_A\mathbf{y}$	$\sigma_{BP}^2 + q_A(\Psi)$
C	$c - 1$	$\mathbf{y}'\mathbf{Q}_C\mathbf{y}$	$\sigma_{BP}^2 + q_C(\Psi)$
A#C	$(a - 1)(c - 1)$	$\mathbf{y}'\mathbf{Q}_{AC}\mathbf{y}$	$\sigma_{BP}^2 + q_{AC}(\Psi)$
Resíduo	$(b - 1)(ac - 1)$	$\mathbf{y}'\mathbf{Q}_{Res}\mathbf{y}$	σ_{BP}^2

Uma revisão sobre o assunto pode ser encontrada em Bailey (2005) e Alcarde (2008).

Como citado anteriormente, as regras acima são descritas para experimentos ortogonais. Quando algum fator longitudinal é incluído no modelo, essa ortogonalidade é quebrada e a obtenção das esperanças dos quadrados médios não mais serão dadas de acordo com tais regras. Entretanto, os diagramas de Hasse continuam sendo uma excelente forma de visualizar a relação existe entre os fatores presentes no experimento.

2.4 Modelos Mistos

A modelagem estatística é baseada na especificação dos valores esperados e das estruturas de variância e covariância. O modelo linear tradicional (fixo), tem como procedimento de estimação o método dos mínimos quadrados ordinários, que é muito restrito a fornecer análises satisfatórias devido às suas pressuposições, tais como independência dos resíduos e homogeneidade de variâncias (BALZARINI, 2002).

Em estudos longitudinais as múltiplas medidas são observadas em cada unidade experimental, em diferentes tempos e, possivelmente, sobre diferentes condições experimentais. Frequentemente, o pesquisador não consegue controlar todas as circunstâncias sobre as quais as medidas são tomadas, sendo comum encontrar conjuntos de dados incompletos e desbalanceados em relação ao tempo, tais conjuntos são difíceis de serem analisados usando o modelo multivariado geral com estrutura de covariância irrestrita, ou não estruturada (LAIRD; WARE, 1982). Nesses estudos, o interesse está na estimação do efeito populacional das covariáveis sobre a variável resposta, dentro e entre indivíduos, bem como

no comportamento individual específico.

Como uma alternativa às duas condições extremas, o modelo linear tradicional e o modelo multivariado, pode-se utilizar os modelos mistos, que são modelos que contém efeitos fixos, além da média geral, e efeitos aleatórios, além do erro. O termo efeito fixo em um modelo descreve o comportamento de toda a população ou daquelas unidades associadas aos níveis repetidos de fatores experimentais. São fatores, geralmente, com poucos níveis, que, usualmente, podem ser controlados pelo pesquisador, enquanto que o termo efeito aleatório descreve a distribuição de um coeficiente dentro da população, são fatores com níveis que estão além do controle do pesquisador (PINHEIRO; BATES, 2000; BALZARINI, 2002).

Tais modelos podem ser vistos como modelos em dois estágios, ou ainda modelos hierárquicos, em que parâmetros de efeito aleatório são amostrados de uma distribuição *a priori* e dados observados são amostrados de distribuições normais independentes, com funções lineares desses efeitos aleatórios e de parâmetros populacionais (de efeito fixo) adicionais, tais como médias.

Com a inclusão de efeitos aleatórios aos fatores, o modelo misto pode acomodar covariâncias entre as observações, sendo assim uma opção para a análise de dados longitudinais. Como consequência dos modelos em dois estágios, a distribuição de probabilidade das múltiplas medidas tem a mesma forma para cada indivíduo, porém os parâmetros dessa distribuição variam com os indivíduos. Tendo escolhido uma família paramétrica como distribuição *a priori*, podem-se realizar inferências sobre os efeitos fixos assim como sobre a distribuição *a priori* (VERBEKE; LESAFFRE, 1997).

A formulação do modelo de dois estágios pode ser descrita como segue. Considere o indivíduo i , e a medida no tempo t_{ij} , com $i = 1, \dots, N$ e $j = 1, \dots, n_i$. Seja $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$, tal que:

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad (1)$$

em que, $\mathbf{Z}_i$ é uma matriz $n_i \times q$, conhecida de covariáveis, $\boldsymbol{\beta}_i$ é um vetor de coeficientes de regressão do indivíduo i e $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$, sendo $\boldsymbol{\Sigma}_i$ a matriz de variâncias e covariâncias entre observações dentro do indivíduo.

O modelo 1 representa o primeiro estágio do modelo misto (FITZMAURICE et al., 2009), em que é ajustado um modelo de regressão para cada indivíduo. O segundo estágio tem como objetivo captar a variabilidade entre indivíduos e é representado por:

$$\boldsymbol{\beta}_i = \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i, \quad (2)$$

tal que, $\mathbf{K}_i$ é uma matriz $q \times p$ de covariáveis conhecidas, $\boldsymbol{\beta}$ é um vetor de parâmetros de regressão desconhecido de ordem p e $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$. De 1 e 2 tem-se:

$$\begin{cases} \mathbf{y}_i &= \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\beta}_i &= \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i \end{cases}.$$

Logo,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad (3)$$

em que, $\mathbf{X}_i = \mathbf{Z}_i \mathbf{K}_i$ é a matriz, $n_i \times p$, dos coeficientes dos efeitos fixos, $\mathbf{Z}_i$ é a matriz dos coeficientes dos efeitos aleatórios, $\boldsymbol{\beta}$ é o vetor de parâmetros de efeito fixo e $\mathbf{b}_i$ é o vetor de parâmetros de efeito aleatório.

O modelo 3 pode ser representado, considerando todos os indivíduos, por:

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

que, matricialmente, é dado por:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \dots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \dots \\ \mathbf{b}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \dots \\ \boldsymbol{\epsilon}_N \end{bmatrix}.$$

Em termos de modelos hierárquicos, o modelo linear misto geral pode ser reescrito de tal modo que:

$$\begin{aligned} \mathbf{Y} | \mathbf{b} &\sim N(\mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b}, \boldsymbol{\Sigma}), \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D}). \end{aligned} \quad (4)$$

Assim, marginalmente, $\mathbf{Y} \sim N(\mathbf{X} \boldsymbol{\beta}, \mathbf{V})$, em que $\mathbf{V} = \mathbf{Z} \mathbf{D} \mathbf{Z}' + \boldsymbol{\Sigma}$. De acordo com Molenberghs e Verbeke (2005), diferentes modelos hierárquicos podem produzir o mesmo modelo marginal, ilustrando que um bom ajuste do modelo marginal pode não ser uma evidência de qual modelo hierárquico foi ajustado. Um tratamento satisfatório do modelo hierárquico somente é possível no contexto bayesiano.

2.5 Estruturas de covariância

A partir do modelo 3, tem-se que:

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}) \text{ e } \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i).$$

As matrizes de variâncias e covariâncias $\mathbf{D}$ e $\boldsymbol{\Sigma}_i$, podem assumir diferentes estruturas, algumas delas apresentadas a seguir, admitindo-se uma situação simples com três ocasiões:

(i) Componentes de variância - CV (“Variance components”):

$$\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

Matriz, geralmente, utilizada como estrutura para $\mathbf{D}$ e $\boldsymbol{\Sigma}_i$, nos modelos ANOVA (todos os fatores são tratados como qualitativos), contendo apenas um parâmetro, σ^2 . Essa matriz supõe independência e homogeneidade de variâncias entre os componentes.

(ii) Simetria composta - SC (“Compound symmetry”):

$$\begin{bmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{bmatrix}$$

A matriz SC contém dois parâmetros e admite homogeneidade de variâncias e covariâncias constantes. Exemplos de utilização dessa matriz podem ser vistos em experimentos casualizados em blocos, para os quais o efeito de blocos é aleatório. Na matriz $\mathbf{V}$, observam-se tantas matrizes SC quantos forem os blocos. Sendo assim, a matriz $\mathbf{V}$ resultante, será dita bloco diagonal.

(iii) Auto regressiva de primeira ordem - AR(1) (“First-order autoregressive model with homogeneous variation”):

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Essa matriz contém dois parâmetros, admite homogeneidade de variâncias e covariâncias decrescentes em função das distâncias entre as observações, ou seja, a correlação decresce à medida que se aumenta a distância entre as observações repetidas, pois $-1 < \rho < 1$. Essa estrutura pode ser utilizada desde que as observações sejam igualmente espaçadas (no tempo ou no espaço).

(iv) Potência (“Power”):

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 \end{bmatrix}$$

Essa matriz também contém dois parâmetros, e apresenta características semelhantes à AR(1), porém seu uso é aconselhado para os casos em que as observações repetidas não são equidistantes.

(v) Componente de variância com heterogeneidade - CV_H (“Variance components with heterogeneous variation”):

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

A matriz CV_H conserva a suposição de independência, porém o mesmo não ocorre com a homogeneidade de variâncias. Essa matriz possui número de parâmetros igual a w , em que w denota a ordem da matriz.

(vi) Banda (“Band”):

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & 0 \\ \sigma_{12}^2 & \sigma_2^2 & \sigma_{23}^2 \\ 0 & \sigma_{23}^2 & \sigma_3^2 \end{bmatrix}$$

A matriz que apresenta covariâncias em bandas supõe heterogeneidade de variâncias e covariâncias presentes considerando-se uma distância pré-estabelecida, em que observações mais distantes que o suposto são independentes.

(vii) Simetria composta com heterogeneidade de variâncias - SC_H (“Compound symmetry with heterogeneous variation”):

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

O número de parâmetros é igual a $w + 1$, em que w representa a ordem da matriz. Admite variâncias e covariâncias distintas, mas correlações iguais.

- (viii) Auto regressiva de ordem um com heterogeneidade de variâncias - $AR_H(1)$ (“First-order autoregressive model with heterogeneous variation”):

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

A matriz $AR_H(1)$ é uma generalização da $AR(1)$, em que são consideradas variâncias e covariâncias distintas entre as observações e correlação que decresce com o tempo.

O número de parâmetros dessa matriz é igual a $w + 1$.

- (ix) Não estruturada - NE (“Unstructured”):

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

A matriz contém o maior número possível de parâmetros, $\frac{w(w+1)}{2}$ parâmetros.

2.6 Estimação dos componentes de variância

Devido às dificuldades advindas da estimação dos parâmetros, em especial dos componentes de variância, a aplicação dos modelos mistos era restrita aos dados longitudinais (FITZMAURICE et al., 2009). Para dados balanceados é comum estimar esses parâmetros pelo método da análise da variância, ou método dos momentos (ANOVA). Uma técnica similar foi desenvolvida por Henderson (1963), para dados desbalanceados.

Os métodos de estimação mais utilizados, atualmente, são os de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML), sendo este último desenvolvido para diminuir o viés em pequenas amostras para máxima verossimilhança. Para que se obtenham estimativas, tanto de máxima verossimilhança quanto de máxima verossimilhança restrita, vários algoritmos têm sido desenvolvidos (FITZMAURICE et al., 2009).

Laird e Ware (1982) mostraram como o algoritmo estimação-maximização (EM) pode ser utilizado para ajustar uma classe de modelos mais gerais para dados longitudinais. Posteriormente, diferentes autores apresentaram novas propostas, algoritmos alternativos, incluindo “Escore de Fisher”, Newton-Raphson e *Average Information REML* (GILMOUR; THOMPSON; CULLIS, 1995).

Para os três métodos citados, as suposições, com relação às variáveis aleatórias, $\mathbf{Y}$ e $\mathbf{b}$, são, conforme apresentadas em 4, dadas por:

$$\begin{aligned}\mathbf{Y}|\mathbf{b} &\sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \boldsymbol{\Sigma}), \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D}).\end{aligned}$$

E marginalmente, $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, em que $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \boldsymbol{\Sigma}$.

2.6.1 Método ANOVA

O método da ANOVA consiste em igualar as esperanças dos quadrados médios de cada fonte de variação, presente na análise da variância, aos seus respectivos quadrados médios.

Em geral, este método é adequado para modelos simples, que envolvem dados balanceados. Os estimadores são não viesados e de variância mínima, são funções de estatísticas suficientes, para as quais podem ser obtidas estimativas dos desvios padrões associados, e uma aproximação dos números de graus de liberdade, por métodos como de Satterthwaite (1946), Fai e Cornelius (1996), Kenward e Roger (1997).

Entretanto, tal método não é indicado para modelos mais complexos, como os de estrutura longitudinal. Quando os dados são não balanceados, não existe um único modo de se obter a tabela da análise da variância, levando a diferentes estimativas para um mesmo componente. Outra característica do método é que não se pode supor que existem componentes de variância diferentes para níveis distintos de alguma determinada fonte de variação.

Devido à construção do método, podem-se obter estimativas negativas para os componentes de variância, fato que torna a propriedade de estimador não viesado pouco interessante. Uma sugestão para contornar tal problema é utilizar a restrição do espaço paramétrico, ou seja, igualar as estimativas negativas a zero. Porém essa solução sacrifica a propriedade do estimador ser não viesado pelo método ANOVA (SEARLE; CASELLA; MCCULLOCH, 1992).

2.6.2 Máxima Verossimilhança

O início da utilização do método da máxima verossimilhança (ML), para estimar componentes de variância, deu-se sob a suposição de normalidade, considerando dados balanceados ou desbalanceados. Para o último caso, métodos iterativos tornaram-se

necessários (SEARLE; CASELLA; MCCULLOCH, 1992).

Como o método ML restringe a estimação ao espaço paramétrico, estimativas negativas dos componentes de variância não podem ser obtidas, uma vez que o espaço paramétrico para tais componentes é estritamente positivo.

De acordo com Harville (1977) os estimadores de ML, são funções de estatísticas suficientes, são consistentes, assintoticamente normais e eficientes. Por outro lado, podem ser viesados, para pequenas amostras, e não consideram a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo. Segundo o mesmo autor, esses estimadores são obtidos sob suposições de uma forma paramétrica particular, geralmente normal, para a distribuição do vetor de dados, porém podem ser “bons” mesmo quando a forma da distribuição dos dados não é especificada.

2.6.3 Máxima Verossimilhança Restrita

Um método alternativo, baseado na verossimilhança, para inferir sobre os componentes de variância nos modelos mistos é o método da máxima verossimilhança restrita (REML).

Nos anos 50, tal método foi desenvolvido para modelos balanceados específicos. Em 1962, foi estendido para todos os modelos balanceados, e o método REML, como tem sido utilizado considerando-se a distribuição normal, foi desenvolvido em 1971 por Patterson e Thompson (HARVILLE, 1977). Em 1990 Verbyla e Cullis aplicaram o método REML em um conjunto de dados longitudinais e, Smyth e Verbyla, em 1996, utilizam a abordagem da verossimilhança condicional para a estimação de REML, generalizando diretamente a interpretação para modelos não normais.

Os estimadores dos componentes de variância, pelo método REML, têm sido amplamente adotados, pois elimina o primeiro dos problemas encontrados no método ML, ou seja, leva em consideração os graus de liberdade envolvidos na estimação dos parâmetros fixos do modelo. Sendo assim, estimativas REML tendem a ser menos viesadas que as estimativas de ML (HARVILLE, 1977 e GILMOUR; THOMPSON; CULLIS, 1995). E, para todos os casos de dados balanceados, as soluções fornecidas pelo método REML a partir dos modelos mistos coincidem com as soluções fornecidas pelo método ANOVA (SEARLE; CASELLA; MCCULLOCH, 1992).

Ao contrário do método ML, que maximiza a função de verossimilhança de todos os contrastes, o método REML maximiza a função de verossimilhança conjunta de

todos os contrastes de erros $\mathbf{Y}^* = \mathbf{L}\mathbf{Y}$, em que $\mathbf{L}$ é uma matriz com $n - \text{posto}(\mathbf{X})$ colunas, de posto completo, com colunas ortogonais as colunas da matriz $\mathbf{X}$, isto é $\mathbf{L}\mathbf{X} = \mathbf{0}$. Em outras palavras, o método REML maximiza a parte da função de verossimilhança que é invariante aos efeitos fixos.

Desse modo, usando uma matriz $\mathbf{L} = [\mathbf{L}_1 \mathbf{L}_2]$, não singular, tal que,

$$\mathbf{L}'_1 \mathbf{X} = \mathbf{I}_p \quad \text{e} \quad \mathbf{L}'_2 \mathbf{X} = \mathbf{0}.$$

Denotando-se $\mathbf{y}^*_j = \mathbf{L}'_j \mathbf{y}$, tal que $j = 1, 2$, tem-se

$$\begin{aligned} \mathbf{y}^*_1 &= \mathbf{L}'_1 \mathbf{y} \\ &= \mathbf{I}_p \boldsymbol{\beta} + \mathbf{L}'_1 \mathbf{Z} \mathbf{b} + \mathbf{L}'_1 \boldsymbol{\epsilon}, \end{aligned}$$

e

$$\begin{aligned} \mathbf{y}^*_2 &= \mathbf{L}'_2 \mathbf{y} \\ &= \mathbf{L}'_2 \mathbf{Z} \mathbf{b} + \mathbf{L}'_2 \boldsymbol{\epsilon}. \end{aligned}$$

Portanto,

$$E(\mathbf{Y}^*) = E \left[\begin{pmatrix} \mathbf{Y}^*_1 \\ \mathbf{Y}^*_2 \end{pmatrix} \right] = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}$$

e

$$\text{Var}(\mathbf{Y}^*) = \text{Var} \left[\begin{pmatrix} \mathbf{Y}^*_1 \\ \mathbf{Y}^*_2 \end{pmatrix} \right] = \begin{pmatrix} \mathbf{L}'_1 \mathbf{V} \mathbf{L}_1 & \mathbf{L}'_1 \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}'_2 \mathbf{V} \mathbf{L}_1 & \mathbf{L}'_2 \mathbf{V} \mathbf{L}_2 \end{pmatrix}.$$

Logo,

$$\begin{pmatrix} \mathbf{Y}^*_1 \\ \mathbf{Y}^*_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{L}'_1 \mathbf{V} \mathbf{L}_1 & \mathbf{L}'_1 \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}'_2 \mathbf{V} \mathbf{L}_1 & \mathbf{L}'_2 \mathbf{V} \mathbf{L}_2 \end{pmatrix} \right].$$

A distribuição completa de $\mathbf{L}'\mathbf{Y}$ pode ser subdividida em uma distribuição condicional, $\mathbf{Y}^*_1 | \mathbf{Y}^*_2$, para a estimação de $\boldsymbol{\beta}$, e uma distribuição marginal, baseada em $\mathbf{Y}^*_2$, para a estimação dos componentes de variância. Esta última é a base da verossimilhança restrita.

Supondo $\boldsymbol{\kappa}' = (\boldsymbol{\gamma}', \boldsymbol{\phi}')$ o vetor de componentes de variância, tal que $\boldsymbol{\gamma}$ contém os componentes de variância associados a $\mathbf{b}$ e $\boldsymbol{\phi}$ os componentes de variância associados a $\boldsymbol{\epsilon}$, sua estimação é baseada no logaritmo da função de verossimilhança restrita,

$$\begin{aligned} l_R &= -\frac{1}{2} \left[\log \det (\mathbf{L}'_2 \mathbf{V}^{-1} \mathbf{L}_2) + \mathbf{y}^{*'}_2 (\mathbf{L}'_2 \mathbf{V} \mathbf{L}_2)^{-1} \mathbf{y} \right] \\ &= -\frac{1}{2} \log \det (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) + \log \det \mathbf{V} + \mathbf{y}' \mathbf{P} \mathbf{y}, \end{aligned}$$

em que,

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1},$$

e ainda,

$$\mathbf{y}'\mathbf{P}\mathbf{y} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

As estimativas REML de κ_l , tal que $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_L)$, são obtidas calculando a função escore, dada por:

$$U(\kappa_l) = \frac{\partial l_R}{\partial \kappa_l} = -\frac{1}{2} \left[\text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \right) - \mathbf{y}'\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P}\mathbf{y} \right], \quad (5)$$

e igualando-a a zero.

Os elementos da matriz informação observada são:

$$\begin{aligned} -\frac{\partial^2 l_R}{\partial \kappa_l \partial \kappa_k} &= \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \right) - \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right) + \\ &+ \mathbf{y}'\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \mathbf{P}\mathbf{y} - \frac{1}{2} \mathbf{y}'\mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \mathbf{P}\mathbf{y}, \end{aligned} \quad (6)$$

e, os elementos da matriz informação esperada são:

$$E \left(-\frac{\partial^2 l_R}{\partial \kappa_l \partial \kappa_k} \right) = \frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right). \quad (7)$$

Entretanto, a solução para $U(\kappa_l) = 0$ exige um algoritmo iterativo. Assim, dada uma estimativa inicial, $\boldsymbol{\kappa}^{(0)}$, uma atualização, $\boldsymbol{\kappa}^{(1)}$, usando-se o algoritmo “Escore de Fisher”, é dada por:

$$\boldsymbol{\kappa}^{(1)} = \boldsymbol{\kappa}^{(0)} + \mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})^{-1} \mathbf{U}(\boldsymbol{\kappa}^{(0)}),$$

em que, $\mathbf{U}(\boldsymbol{\kappa}^{(0)})$ é o vetor escore, apresentado na equação 5, e $\mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})$ representa a matriz de informação esperada de $\boldsymbol{\kappa}$, apresentada na equação 7, avaliada em $\boldsymbol{\kappa}^{(0)}$.

Para modelos de grandes dimensões ou grandes conjuntos de dados, a avaliação dos termos de traço nas equações 6 e 7 pode não ser viável ou computacionalmente intensiva. Nesses casos será utilizado o algoritmo AI (GILMOUR; THOMPSON; CULLIS, 1995), que de acordo com os autores apresenta propriedades de convergência semelhantes ao algoritmo “Escore de Fisher”, mas evita a sobrecarga computacional. Assim como o algoritmo “Escore de Fisher”, pode ser usado para obter estimativas negativas de componentes de variância, removendo as restrições do espaço paramétrico.

A essência do algoritmo AI é substituir a matriz de informação esperada, $\mathbf{I}(\kappa_l, \kappa_k)$, por $\mathcal{I}_{\mathcal{A}}(\kappa_l, \kappa_k)$, em que seus elementos são descritos por:

$$\mathcal{I}_{\mathcal{A}}(\kappa_l, \kappa_k) = \frac{1}{2} \mathbf{y}'\mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \mathbf{P}\mathbf{y}.$$

A matriz $\mathcal{I}_{\mathcal{A}}$ é obtida pela média dos termos presentes nas equações 6 e 7, e aproximando $\mathbf{y}'\mathbf{P}\frac{\partial^2\mathbf{V}}{\partial\kappa_l\partial\kappa_k}\mathbf{P}\mathbf{y}$ por sua esperança, $\text{tr}\left(\mathbf{P}\frac{\partial^2\mathbf{V}}{\partial\kappa_l\partial\kappa_k}\right)$, nos casos em que $\frac{\partial^2\mathbf{V}}{\partial\kappa_l\partial\kappa_k} \neq 0$. Para modelos ANOVA, os termos em $\mathcal{I}_{\mathcal{A}}$ são médias exatas dos termos presentes nas equações 6 e 7.

2.7 Estimação e predição

Os procedimentos mais utilizados para a obtenção do estimador do vetor de parâmetros de efeito fixo, e preditor do vetor de parâmetros de efeito aleatório, são o BLUE (*best linear unbiased estimator* - melhor estimador não viesado) e o BLUP (*best linear unbiased predictor* - melhor preditor não viesado), respectivamente. Esses são denominados “melhores” no sentido que minimizam a variância amostral, “linear”, pois são funções lineares de $\mathbf{y}$ e “não viesados”, porque $E(\text{BLUE}(\boldsymbol{\beta})) = \boldsymbol{\beta}$ e $E(\text{BLUP}(\mathbf{b})) = \mathbf{b}$ (WALSH, 2010).

2.7.1 Estimação dos termos fixos

São dois os casos de estimação a serem considerados seguindo a suposição de normalidade dos vetores $\mathbf{Y}$ e $\mathbf{b}$: o primeiro, **i)**, em que se toma como conhecido o vetor de parâmetros de componentes de variância, $\boldsymbol{\kappa}' = (\boldsymbol{\gamma}', \boldsymbol{\phi}')$, e o segundo, **ii)**, trata-se do caso comumente encontrado, em que se assume $\boldsymbol{\kappa}$ desconhecido.

i) Caso especial: assumindo-se $\boldsymbol{\kappa}$ conhecido

Considerando conhecidos os componentes de variância do modelo, os parâmetros fixos podem ser obtidos pelo método dos mínimos quadrados generalizados e

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad (8)$$

tal que, $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1})$ e é o BLUE de $\boldsymbol{\beta}$.

ii) Caso geral: parâmetro $\boldsymbol{\kappa}$ é desconhecido

Para os casos em que os componentes de variância são desconhecidos, suas estimativas obtidas pelos métodos mencionados podem ser utilizadas, de modo que a matriz $\hat{\mathbf{V}}$, seja dada por:

$$\hat{\mathbf{V}} = \mathbf{Z}\hat{\mathbf{D}}\mathbf{Z}' + \hat{\boldsymbol{\Sigma}},$$

e, a estimativa de $\boldsymbol{\beta}$, em função da estimativa de $\boldsymbol{\kappa}$, é dada por:

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

em que, $\text{Var}(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\kappa}})) = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$.

2.7.2 Predição dos termos aleatórios

Henderson (1963) mostrou que $\hat{\mathbf{b}}$, dado por:

$$\hat{\mathbf{b}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad (9)$$

é o BLUP de $\mathbf{b}$, que é equivalente a $E(\mathbf{b}|\mathbf{Y})$, sob a suposição de normalidade multivariada.

2.7.3 Estimação dos termos fixos e predição dos termos aleatórios simultaneamente

Henderson (1953) obteve as seguintes equações de modelos mistos maximizando a função de verossimilhança conjunta de $\mathbf{Y}$ e $\mathbf{b}$, em relação aos vetores $\boldsymbol{\beta}$ e $\mathbf{b}$,

$$\begin{bmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{bmatrix},$$

a partir das quais é possível estimar os efeitos fixos e prever os efeitos aleatórios, simultaneamente, sem a necessidade de calcular a matriz $\mathbf{V}^{-1}$.

A estimativa de $\boldsymbol{\beta}$ e a predição de $\mathbf{b}$ são dadas, respectivamente, por:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left\{ \mathbf{X}' \left[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1} \right] \mathbf{X} \right\}^{-1} \times \\ &\quad \times \mathbf{X}' \left[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1} \right] \mathbf{y} \end{aligned} \quad (10)$$

e

$$\hat{\mathbf{b}} = \left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \right)^{-1} \mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (11)$$

O estimador 10 é equivalente ao estimador 8, que é o melhor estimador linear não viesado de $\boldsymbol{\beta}$, e, o preditor 11 é equivalente ao preditor 9, que é o melhor preditor linear não viesado de $\mathbf{b}$.

Considerando-se os modelos mistos descritos em termos de modelos hierárquicos, apresentados em 4, tem-se que a distribuição conjunta de $\mathbf{Y}$ e $\mathbf{b}$ condicional

à κ é dada por:

$$\begin{aligned}
 f(\mathbf{Y}, \mathbf{b}|\kappa) &= f(\mathbf{Y}|\mathbf{b}, \kappa) \times f(\mathbf{b}|\kappa) \\
 &= (2\pi)^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb})' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb}) \right\} \times \\
 &\quad \times (2\pi)^{-\frac{1}{2}} |\mathbf{D}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{b}' \mathbf{D}^{-1} \mathbf{b} \right\} \\
 &= (2\pi)^{-1} |\Sigma|^{-\frac{1}{2}} |\mathbf{D}|^{-\frac{1}{2}} \times \\
 &\quad \times \exp \left\{ -\frac{1}{2} [(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb})' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb}) + \mathbf{b}' \mathbf{D}^{-1} \mathbf{b}] \right\}.
 \end{aligned} \tag{12}$$

Logo, o logaritmo da distribuição conjunta (equação 12) é proporcional a l dado por:

$$l \propto \log |\Sigma| + \log |\mathbf{D}| + (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb})' \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta - \mathbf{Zb}) + \mathbf{b}' \mathbf{D}^{-1} \mathbf{b},$$

e suas derivadas parciais, com relação aos parâmetros β e $\mathbf{b}$, podem ser observadas na equação a seguir:

$$\begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \Sigma^{-1} \mathbf{y} - \mathbf{X}' \Sigma^{-1} \mathbf{X} \beta - \mathbf{X}' \Sigma^{-1} \mathbf{Z} \mathbf{b} \\ \mathbf{Z}' \Sigma^{-1} \mathbf{y} - \mathbf{Z}' \Sigma^{-1} \mathbf{X} \beta - \mathbf{Z}' \Sigma^{-1} \mathbf{Z} \mathbf{b} - \mathbf{D}^{-1} \mathbf{b} \end{bmatrix}. \tag{13}$$

Quando as equações apresentadas em 13 são igualadas a zero, são obtidas as equações dos modelos mistos (HENDERSON, 1953), dadas por:

$$\begin{bmatrix} \mathbf{X}' \Sigma^{-1} \mathbf{X} & \mathbf{X}' \Sigma^{-1} \mathbf{Z} \\ \mathbf{Z}' \Sigma^{-1} \mathbf{X} & \mathbf{Z}' \Sigma^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \Sigma^{-1} \mathbf{y} \\ \mathbf{Z}' \Sigma^{-1} \mathbf{y} \end{bmatrix}. \tag{14}$$

A estimativa de β e a predição de $\mathbf{b}$ são dadas, respectivamente, por:

$$\begin{aligned}
 \hat{\beta} &= \left\{ \mathbf{X}' \left[\Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}' \Sigma^{-1} \mathbf{Z} + \mathbf{D}^{-1})^{-1} \mathbf{Z}' \Sigma^{-1} \right] \mathbf{X} \right\}^{-1} \times \\
 &\quad \times \mathbf{X}' \left[\Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}' \Sigma^{-1} \mathbf{Z} + \mathbf{D}^{-1})^{-1} \mathbf{Z}' \Sigma^{-1} \right] \mathbf{y}
 \end{aligned}$$

e,

$$\hat{\mathbf{b}} = (\mathbf{Z}' \Sigma^{-1} \mathbf{Z} + \mathbf{D}^{-1})^{-1} \mathbf{Z}' \Sigma^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}),$$

que exigem soluções para $\kappa' = (\gamma', \phi')$. Na prática, substituem-se γ e ϕ por suas estimativas REML, $\hat{\gamma}$ e $\hat{\phi}$.

2.8 Suavização spline cúbica

Considerando-se um conjunto de dados que, por meio de uma análise visual ou de conhecimentos prévios a respeito do comportamento da variável resposta em função dos níveis do fator longitudinal, pode-se supor que tal comportamento seja não linear, um método proposto por Verbyla et al. (1999) para sua modelagem é a utilização das suavizações splines cúbicas, que foram introduzidos na estatística no contexto de ajustar uma suave, mas desconhecida função aos dados na presença de um erro, ou seja, uma curva aproximada.

Nesta seção serão abordados os modelos de suavização spline cubica conforme descrito por Verbyla et al. (1999), apresentando uma breve introdução, considerando-se a observação de um único indivíduo, com o qual pode-se notar que os modelos de suavização splines cúbicos podem ser entendidos como modelos mistos, em seguida a observação de vários indivíduos e por fim a observação de indivíduos dispostos em um experimento conforme um delineamento.

2.8.1 Suavização Spline cúbica: um único indivíduo

Como ilustração dos conceitos básicos do método proposto, sejam as observações y_i , $i = 1, 2, \dots, p$ nos valores t_i de uma variável explanatória T, tais que os valores y_i satisfazem a seguinte relação:

$$y_i = g(t_i) + e_i,$$

em que $t_1 < t_2 < \dots < t_p$, $g(\cdot)$ é uma função suave, porém desconhecida e os erros, e_i , são normais com média zero e matriz de covariâncias $\Sigma = \sigma^2 \mathbf{R}$, com componentes de variância desconhecidos. Utilizando-se a notação matricial, tem-se:

$$\mathbf{y} = \mathbf{g} + \mathbf{e}.$$

A função $\mathbf{g}(\cdot)$, e portanto $\mathbf{g}$, é escolhida de tal modo a maximizar o logaritmo da função de verossimilhança penalizada, dada por:

$$\begin{aligned} l &= -\frac{1}{2} \log\{\det(\Sigma)\} - \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^p \sum_{j=1}^p R^{ij} [y_i - g(t_i)][y_j - g(t_j)] + \lambda_s \int g''(t)^2 dt \right\} \\ &= -\frac{1}{2} \log\{\det(\Sigma)\} - \frac{1}{2\sigma^2} \left[(\mathbf{y} - \mathbf{g})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{g}) + \lambda_s \int g''(t)^2 dt \right], \end{aligned}$$

em que R^{ij} são elementos de $\mathbf{R}^{-1}$. A solução é uma suavização spline cúbica, e o parâmetro λ_s , governa o quão suave é a curva.

Seja $h_j = t_{j+1} - t_j$, $j = 1, 2, \dots, p-1$. Defina-se Δ e $\mathbf{G}_s$, $p \times (p-2)$ e $(p-2) \times (p-2)$, respectivamente, matrizes em bandas, tais que seus elementos diferentes de zero são dados por:

$$\Delta_{ii} = \frac{1}{h_i}, \quad \Delta_{i+1,j} = -\left(\frac{1}{h_i} + \frac{1}{h_{i+1}}\right), \quad \Delta_{i+2,i} = \frac{1}{h_{i+1}},$$

$$G_{s;i,i+1} = G_{s;i+1,i} = \frac{h_{i+1}}{6}, \quad G_{s,ii} = \frac{h_i + h_{i+1}}{3}.$$

Se $\mathbf{K} = \Delta \mathbf{G}_s^{-1} \Delta'$, então nos pontos de delineamento da suavização spline, as estimativas de $\mathbf{g}$ são dadas por:

$$\hat{\mathbf{g}} = (\Sigma^{-1} + \lambda_s \mathbf{K})^{-1} \Sigma^{-1} \mathbf{y}.$$

Seja $\mathbf{X}_s$ uma matriz $p \times 2$, cujas colunas são um vetor de uns, $\mathbf{1}_p$, o vetor $\mathbf{t}' = (t_1, t_2, \dots, t_p)'$, e seja $\mathbf{Z}_s = \Delta(\Delta' \Delta)^{-1}$. Defina-se $\mathbf{V} = \sigma^2(\mathbf{R} + \lambda_s^{-1} \mathbf{Z}_s \mathbf{G}_s \mathbf{Z}_s')$. Desse modo,

$$\hat{\mathbf{g}} = \mathbf{X}_s \hat{\boldsymbol{\beta}}_s + \mathbf{Z}_s \hat{\mathbf{u}}_s,$$

em que,

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

$$\hat{\mathbf{u}}_s = (\mathbf{Z}_s' \Sigma^{-1} \mathbf{Z}_s + \lambda_s \mathbf{G}_s^{-1})^{-1} \mathbf{Z}_s' \Sigma^{-1} (\mathbf{y} - \mathbf{X}_s \hat{\boldsymbol{\beta}}_s).$$

com $i = 1, 2, \dots, p-2$.

Esse resultado mostra que a suavização spline pode ser escrita como um modelo misto, tal que:

$$\mathbf{y} = \mathbf{X}_s \boldsymbol{\beta}_s + \mathbf{Z}_s \mathbf{u}_s + \mathbf{e}, \quad (15)$$

em que, $\mathbf{u}_s$ é um vetor $(p-2) \times 1$ e $\mathbf{u}_s \sim N(\mathbf{0}, \sigma_s^2 \mathbf{G}_s)$ e o parâmetro de suavização é uma razão de variâncias, $\lambda_s = \frac{\sigma^2}{\sigma_s^2} = \gamma_s^{-1}$.

2.8.2 Suavização Spline cúbica: vários indivíduos

Uma das pré-suposições da experimentação é a repetição, sendo assim, frequentemente os valores de uma variável explanatória não são únicos ao longo das unidades experimentais. Os resultados obtidos na subseção anterior, usando-se modelos mistos (modelo 15), permanecem válidos quando consideradas matrizes de delineamento contendo linha

duplicadas. O que significa prmultiplicar as matrizes $\mathbf{X}_s$ e $\mathbf{Z}_s$ por uma matriz de delimitamento binria, $\mathbf{N}_g$, em que a i -sima coluna contm n_i unidades, correspondendo ao nmero de repeties do valor t_i . Se $\mathbf{y}$ um vetor de observaes $n \times 1$, o modelo pode ser reescrito, tal que

$$\mathbf{y} = \mathbf{N}_g(\mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\mathbf{u}_s) + \mathbf{e}. \quad (16)$$

A suavizao ento ajustada para as mdias dos valores para as repeties, com uma possvel ponderao, quando considerados nmeros diferentes de repeties para os nveis da varivel T.

O modelo 16 pode ser considerado sob a suposio de que as tendncias so suaves, entretanto, podem ocorrer desvios nas tendncias, e na presena de efeitos no suaves, ou seja, desses desvios, um ajuste utilizando o mtodo proposto, incorpora os componentes suaves e no suaves.

Se existem repeties no valor t_i , os efeitos no suaves podem ser modelados, e o modelo passa a ser expresso por:

$$\mathbf{y} = \mathbf{N}_g(\mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\mathbf{u}_s + \mathbf{u}_g) + \mathbf{e},$$

em que $\mathbf{u}_g$ o efeito aleatrio padro para T, tomado como um fator qualitativo, tal que $\mathbf{u}_g \sim N(\mathbf{0}, \sigma_g^2 \mathbf{I}_p)$.

Esse tipo de termo permite que a tendncia aleatria seja decomposta nos componentes suave e no suave. Se a tendncia no suave no estiver presente, o componente de varincia σ_g^2 ser estimado prximo de zero. Vale ressaltar que existem muitas outras possveis especificaes para $\mathbf{u}_g$.

2.8.3 Suavizao Spline cbica: presena de fatores qualitativos

Considerando-se um novo exemplo, contendo dois fatores cruzados, um fator de tratamento qualitativo A e um fator quantitativo T. Seja y_{ijk} a varivel resposta avaliada no i -simo valor do fator T, no j -simo nvel do fator A e na k -sima repetio da combinao (i, j) , $(k = 1, 2, \dots, n_{ij})$, com $n = \sum_{i=1}^p \sum_{j=1}^r n_{ij}$.

Seja o modelo saturado,

$$y_{ijk} = \eta_{ijk} + e_{ijk},$$

em que,

$$\begin{aligned}\eta_{ijk} &= \mu + \tau_i + \alpha_j + \zeta_{ij} \\ &= \mu + \tau(t_i) + \alpha_j + \zeta_j(t_i),\end{aligned}\tag{17}$$

com μ , τ_i , α_j e ζ_{ij} representando respectivamente, os efeitos da média geral, do fator T, do fator A e da interação entre os fatores T e A. Os mesmos efeitos são reescritos explicitamente em função de t_i .

Para a notação matricial, tomando-se todos os níveis do fator T, tem-se:

$$\boldsymbol{\eta}_{jk} = \mathbf{1}_p \mu + \boldsymbol{\tau} + \mathbf{1}_p \alpha_j + \boldsymbol{\zeta}_j,$$

em que $\boldsymbol{\tau}$ e $\boldsymbol{\zeta}_j$ são vetores de ordem $p \times 1$ e $\mathbf{1}_p$ é um vetor de uns de ordem p .

Em termos dos modelos mistos, tal como no modelo 15, $\boldsymbol{\tau}$ e $\boldsymbol{\zeta}_j$ podem ser expressos por:

$$\boldsymbol{\tau} = \mathbf{X}_s \boldsymbol{\beta}_{s,0} + \mathbf{Z}_s \mathbf{u}_{s,0} + \mathbf{u}_{g,0}\tag{18}$$

$$\boldsymbol{\zeta}_j = \mathbf{X}_s \boldsymbol{\beta}_{s,j} + \mathbf{Z}_s \mathbf{u}_{s,j} + \mathbf{u}_{g,j},\tag{19}$$

em que, $\mathbf{u}_{s,0} \sim N(\mathbf{0}, \sigma_{s,0}^2 \mathbf{G}_s)$ e $\mathbf{u}_{s,j} \sim N(\mathbf{0}, \sigma_{s,j}^2 \mathbf{G}_s)$, com $j = 1, 2, \dots, r$. E, $\mathbf{u}_{g,0} \sim N(\mathbf{0}, \sigma_{g,0}^2 \mathbf{I}_p)$ e $\mathbf{u}_{g,j} \sim N(\mathbf{0}, \sigma_{g,j}^2 \mathbf{I}_p)$, com $j = 1, 2, \dots, r$. Estruturas de variâncias mais complexas podem ser utilizadas.

2.8.4 Representação da análise da variância

Considere A o efeito do fator A, T o efeito do fator T e, $A\#T$ o efeito da interação entre os fatores A e T. O modelo baseado na equação 17 apresenta a seguinte expansão da fórmula estrutural:

$$y \sim A + T + A\#T,$$

em que, o modelo misto completo, incorporando os modelos 18 e 19, é dado pela seguinte expansão:

$$y \sim A + \text{lin}(t) + \text{spl}(t) + \text{dev}(t) + A\#\text{lin}(t) + A\#\text{spl}(t) + A\#\text{dev}(t),$$

com $\text{lin}(t)$ denotando o efeito linear do fator T tomado como uma variável contínua com valores t_i , que contribui com um único parâmetro (coeficiente angular). O termo $\text{spl}(t)$

denota o correspondente componente aleatório spline ($\mathbf{N}_g \mathbf{Z}_s \mathbf{u}_s$), e $\text{dev}(t)$ denota o efeito aleatório do fator T ($\mathbf{N}_g \mathbf{u}_g$).

Tal modelo pode ser resumido na tabela da análise da variância apresentada a seguir:

Tabela 4 - Decomposição da tabela da análise da variância, considerando um experimento fatorial, em que o fator A é qualitativo e o fator T quantitativo

Fontes de Variação	Tipo	gl ou ne
A	Fixo	$r - 1$
T		
$\text{lin}(t)$	Fixo	1
$\text{spl}(t)$	Aleatório	$p - 2$
$\text{dev}(t)$	Aleatório	p
$A \# T$		
$A \# \text{lin}(t)$	Fixo	$r - 1$
$A \# \text{spl}(t)$	Aleatório	$r(p - 2)$
$A \# \text{dev}(t)$	Aleatório	rp
Resíduo	Aleatório	

gl denotam o número de graus de liberdade para efeitos fixos e ne o número de efeitos para efeitos aleatórios

2.8.5 Suavização Spline cúbica: dados longitudinais

Quando se trabalha com dados longitudinais, a variável quantitativa T é geralmente o tempo, podendo ser também profundidades, entre outros fatores de interesse. Para simplificar, será assumido que todas as unidades foram medidas nos mesmos tempos, ou seja, o conjunto de dados em questão é balanceado e balanceado em relação ao tempo. Seja $y_{jk}(t_i)$ a variável resposta no i -ésimo tempo, para a k -ésima unidade que recebeu o j -ésimo nível do fator A, logo o modelo saturado é dado por:

$$y_{jk}(t_i) = \eta_{ijk} + e_{ijk},$$

em que,

$$\eta_{ijk} = \mu(t_i) + \alpha_j(t_i) + \zeta_{jk}(t_i).$$

O vetor de médias para a k -ésima unidade que recebeu o j -ésimo nível do fator A é dado por:

$$\eta_{jk} = \mu + \alpha_j + \zeta_{jk},$$

sendo cada vetor uma função do tempo.

Assim, o modelo completo pode ser expresso por:

$$y \sim \text{lin}(t) + \text{spl}(t) + \text{dev}(t) + A + A\#\text{lin}(t) + A\#\text{spl}(t) + A\#\text{dev}(t) + \\ + \text{unidade} + \text{unidade}\#\text{lin}(t) + \text{unidade}\#\text{spl}(t).$$

Os termos unidade e unidade#lin(t) são os interceptos e os coeficientes angulares para os indivíduos, tomados como de efeito aleatório. O termo unidade#spl(t) é um componente aleatório spline para unidade. Tais termos aumentam a flexibilidade da estrutura de covariância. O termo unidade#dev(t) corresponde ao termo de erro, por isso não foi representado no modelo.

2.9 Construção de modelos mistos

Piepho et al. (2003) apresentaram um conjunto de regras para construção de modelos mistos considerando modelos para fatores de tratamento qualitativos e dados não longitudinais, denominados modelos do tipo ANOVA. Embora inicialmente os autores discriminem fatores não aleatorizados e aleatorizados, nos modelos finais, tais diferenças não se apresentam tão claras, pois os fatores que indexam as unidades, (não aleatorizados) são substituídos pelos fatores aleatorizados (de tratamento), confundidos com os mesmos, é o caso em que o efeito do acaso é substituído pelo efeito da interação entre Blocos e Tratamentos, em um experimento instalado de acordo com um delineamento casualizado em blocos.

Piepho et al. (2004) aprimoraram essas técnicas e desenvolveram um novo conjunto de regras para experimentos longitudinais, com o inconveniente de substituir fatores não aleatorizados por fatores aleatorizados. No modelo final porém, fatores longitudinais são tratados como fatores não aleatorizados.

Como uma proposta de modificação desses dois trabalhos e, como extensão do método proposto por Brien e Bailey (2006), Brien e Demétrio (2009) descreveram a formulação de modelos mistos, incluindo experimentos longitudinais, também por um conjunto de regras, porém discriminaram, inclusive no modelo final, os fatores não aleatorizados dos fatores aleatorizados, e afirmaram que os fatores longitudinais não são fatores não

aleatorizados, apenas nenhum fator é aleatorizado a eles, salientando assim a importância da construção de modelos baseados na aleatorização.

O objetivo do método proposto por Brien e Demétrio (2009), também denominado por processo em três estágios, é obter de forma simbólica o modelo fixo, bem como o modelo aleatório, que consistem dos termos correspondentes, respectivamente, às submatrizes da matriz $\mathbf{X}$, e às submatrizes de $\mathbf{Z}$, em que $\mathbf{X}$ e $\mathbf{Z}$ são matrizes no modelo descrito pela equação 3. Também há interesse em se obter os termos que indexam as unidades experimentais, correspondentes à matriz Σ , que é a matriz de variâncias e covariância de ϵ .

A construção do modelo misto será apresentada a seguir e como ilustração para as técnicas, seja o exemplo apresentado no início desta Seção, no qual considerou-se um experimento casualizado em blocos, com b o número de blocos, e esquema de tratamento fatorial, em que o fator qualitativo A possui a níveis e o fator qualitativo C, c níveis. Considerando-se agora que cada uma das abc parcelas foram observadas ao longo de t tempos.

1. Determinar fórmulas para o modelo intra-camada composto de um modelo intra-camada aleatório (IR) e um modelo intra-camada fixo (IF):

- (a) Identificar os conjuntos de objetos e indicar a unidade observacional, isto é, a menor unidade a partir da qual um único valor da variável resposta é obtido. Para os experimentos com duas camadas existirão dois conjuntos de objetos: o não aleatorizado e o aleatorizado.

No exemplo: Os conjuntos de objetos são: **i)** $abct$ combinações tempos-unidades, e **ii)** ac tratamentos; a unidade observacional é a combinação tempo-parcela.

- (b) Determinar as duas camadas: a camada para os fatores não aleatorizados e a camada para os fatores aleatorizados. Uma camada é composta por fatores no painel de um diagrama de aleatorização (Brien; Bailey, 2006).

No exemplo : A camada para os fatores não aleatorizados é {Blocos, Parcelas}, e a camada para os fatores aleatorizados é {A, C}.

- (c) Identificar os fatores longitudinais, se existirem.

No exemplo: O fator longitudinal é Tempos.

- (d) Determinar as fórmulas intra-camada especificando as relações de aninhamento, de cruzamento e aditivas entre os fatores em cada camada. Esse relacionamento

dependerá das relações intrínsecas e daquelas impostas no delineamento. Usualmente, fatores longitudinais são cruzados com outros fatores dentro de uma camada e quando possível devem ser alocados do lado direito de outros. A fórmula derivada a partir dos fatores não aleatorizados especifica o modelo aleatório sob aleatorização, ou modelo IR. O modelo marginal implícito é equivalente ao modelo de aleatorização, exceto que o último permite componentes de variância negativos.

No exemplo: IR: (Blocos/Parcelas)* Tempos

IF: A*C

2. Converter o modelo de termos intra-camada composto de um modelo aleatório homogêneo (HR) e um modelo fixo (F) da seguinte forma:

- (a) Adicionar ao modelo IF termos intra-camada que são de interesse/importância para o pesquisador, isto é, fatores não aleatorizados, como Ambiente ou Sexo, e suas interações com fatores aleatorizados, ou, interação com Provedor em experimentos de avaliação sensorial. Se existem fatores longitudinais, então, usualmente, haverá interação com fatores aleatorizados e fatores não aleatorizados de interesse do pesquisador, tais interações devem também ser adicionadas ao modelo IF.

No exemplo: Existe interesse em avaliar o comportamento das parcelas, que receberam determinado tratamento, ao longo do tempo. Logo, o modelo fixo passa a ser F: A*C*Tempos

- (b) Aumentar os modelos IR e IF com termos potencialmente importantes não considerados no delineamento do experimento. Isso pode envolver fatores identificados como importantes fontes de variação somente após o início do experimento, ou covariáveis contínuas que não puderam ser controladas no delineamento.

No exemplo: Não há necessidade de inclusão de termos no modelo.

- (c) Decidir quais termos são de efeito fixo e quais termos são de efeito aleatório. Isso pode ser feito considerando como termos fixos aqueles que envolvem somente fatores fixos e, caso contrário, o termo é aleatório.

No exemplo: Os fatores de efeito fixo são A, C e Tempos, e os fatores de efeito aleatório são Blocos e Parcelas.

- (d) Obter os modelos finais HR e F deslocando os termos que são fixos no modelo IR, previamente modificado, para o modelo F final e, os termos que são aleatórios no modelo IF, previamente modificado, para o modelo HR final.

No exemplo: As estruturas IR e F são dadas por:

$$\text{IR: } \text{Blocos} + \text{Blocos} \wedge \text{Parcelas} + \text{Tempos} + \text{Blocos} \wedge \text{Tempos} + \\ + \text{Blocos} \wedge \text{Parcelas} \wedge \text{Tempos},$$

$$\text{F: } A + C + A \wedge C + \text{Tempos} + A \wedge \text{Tempos} + C \wedge \text{Tempos} + A \wedge C \wedge \text{Tempos}.$$

Após as modificações necessárias, os modelos HR e F são:

$$\text{HR: } \text{Blocos} + \text{Blocos} \wedge \text{Parcelas} + \text{Blocos} \wedge \text{Tempos} + \text{Blocos} \wedge \text{Parcelas} \wedge \text{Tempos}, \\ = \text{Blocos} / \text{Tempos} + \text{Blocos} \wedge \text{Parcelas} / \text{Tempos}$$

$$\text{F: } A + C + A \wedge C + \text{Tempos} + A \wedge \text{Tempos} + C \wedge \text{Tempos} + A \wedge C \wedge \text{Tempos}. \\ = A * C * \text{Tempos}.$$

3. Formular o modelo misto geral, obtendo os modelos gerais aleatório (GR) e fixo (GF), a partir dos modelos HR e F:

- (a) Se existirem fatores longitudinais, permitir correlações desiguais entre seus níveis, assegurando que todo termo de erro longitudinal tenha sido incluído no modelo aleatório. Termos de erro longitudinal são termos aleatórios da forma “(indivíduo)^(fatores longitudinais)”. Um indivíduo, para um fator longitudinal, é um fator generalizado cujos níveis são unidades nas quais sucessivas observações são realizadas. Os indivíduos são geralmente compostos por fatores não aleatorizados que não são longitudinais, mas em algumas ocasiões incluirão fatores aleatorizados e outros fatores longitudinais. Se forem considerados modelos envolvendo correlações desiguais para os seus termos de erro longitudinal, os indivíduos serão escritos em *itálico*.

O operador “gf” é aplicado à parte da formula envolvendo fatores longitudinais, de modo que nenhuma restrição é considerada na forma da correlação entre os níveis das combinações dos fatores longitudinais. Então, aplica-se a esse fator generalizado o operador “uc”. Se o fator longitudinal for aleatorizado, será necessário incorporar seus termos de erro longitudinal adicionando interações entre os mesmos e o indivíduo. Um fator longitudinal pode ter vários termos como indivíduos. Se estruturas de correlações desiguais são restritas aos termos envolvendo somente fatores aleatorizados, termos de erro longitudinal devem ser restritos aos termos compostos somente por tais fatores.

No exemplo: O indivíduo é $\text{Blocos} \wedge \text{Parcelas}$, e é esperado que existam correlações distintas entre observações em tempos diferentes. Logo, o modelo geral inicial é dado

por:

$$\begin{aligned} \text{GR: } & \text{Blocos}/\text{gf}(\text{Tempos}) + (\text{Blocos}\wedge\text{Parcelas})/\text{gf}(\text{Tempos}) \\ & \Rightarrow \text{Blocos}/\text{uc}(\text{Tempos}) + (\text{Blocos}\wedge\text{Parcelas})/\text{uc}(\text{Tempos}). \end{aligned}$$

- (b) Fazer outras mudanças no modelo GR, tais como permitir níveis de variâncias desiguais entre os níveis de fatores longitudinais ou correlações desiguais entre os níveis de outros fatores aleatórios.

No exemplo: Não são necessárias modificações no modelo GR.

- (c) Considerar se um termo aleatório precisa ser removido do modelo GR para evitar superparametrização.

No exemplo: Não é necessário no modelo.

- (d) Se apropriado, aplicar o operador “td” para especificar outras parametrizações para os termos fixos, tais como tendências polinomiais, suavizadas ou não lineares.

No exemplo: Tendências sistemáticas em Tempos são de interesse, mas não nos fatores qualitativos A e C. Assim, o modelo GF é dado conforme segue:

$$\text{GF: } A * C * \text{td}(\text{Tempos}).$$

Na Tabela 5 são apresentados os operadores de tendência e os operadores para correlações desiguais, os quais serão aplicados aos fatores generalizados. Os operadores de tendências, $\text{td}(\cdot)$, trabalham com suposições a respeito da parte fixa do modelo, ou seja, sobre a forma da matriz $\mathbf{X}$, enquanto que os operadores de correlações desiguais, $\text{uc}(\cdot)$, alteram as formas das matrizes $\mathbf{D}$ e $\mathbf{\Sigma}$.

Tabela 5 - Resumo para os operadores de tendência e de correlação desigual

Operadores	
de tendência	
$td(\cdot)$	Especifica o modelo de tendência na resposta, como uma função de valores das variáveis quantitativas associadas ao fator
$lin(\cdot)$	Especifica tendência linear nos fatores quantitativos
$pol(\cdot)$	Especifica tendência polinomial nos fatores quantitativos
$spl(\cdot)$	Especifica que as tendências estão sendo modeladas usando suavização spline cúbica, ajustando-se um componente de variância ao modelo aleatório
$dev(\cdot)$	Especifica desvios aleatórios em torno da tendência ajustada
da matriz de correlação	
$uc(\cdot)$	Operador de correlação desigual especifica que, nos modelos aleatórios, serão consideradas diferentes correlações entre diferentes níveis dos fatores generalizados
$id(\cdot)$	Operador matriz de correlação identidade (CV)
$cor(\cdot)$	Operador matriz de correlação uniforme (SC)
$corg(\cdot)$	Operador matriz de correlação geral
$diag(\cdot)$	Operador matriz de correlação diagonal (CV_H)
$ar1(\cdot)$	Operador de correlação autorregressiva de ordem 1
$corb(\cdot)$	Operador matriz de correlação em bandas
$us(\cdot)$	Operador matriz de correlação irrestrito (NE)
de variâncias desiguais	
Adicionar “h” ao nome do operador matriz de correlação	

Fonte: Brien e Demétrio (2009)

2.10 Seleção de modelos

Na seleção de um modelo busca-se um modelo parcimonioso, em termos do número de parâmetros, e que ao mesmo tempo seja o que melhor prediz a variável resposta.

Na seleção de modelos para a análise de um conjunto de dados, consideram-se pesquisas objetivas, a amostragem e o delineamento do estudo, os conhecimentos prévios

sobre preditores importantes, e características do indivíduo. Usam-se ferramentas analíticas, tais como testes de hipóteses e critérios de informação, que são descritos, a seguir.

2.10.1 Teste da razão de verossimilhanças

Os testes de razão de verossimilhanças pertencem a uma classe de testes baseados na comparação dos valores das funções de verossimilhança para dois modelos aninhados. Um modelo mais simples (modelo aninhado), referente à hipótese nula, é denominado aninhado a um modelo mais geral (modelo de referência), quando o espaço paramétrico do primeiro modelo é um subespaço do segundo, ou seja, quando os parâmetros do modelo aninhado podem ser obtidos impondo-se certas restrições aos parâmetros do modelo de referência (WEST; WELCH; GALECKI, 2007).

No contexto dos modelos lineares mistos, o teste da razão de verossimilhanças pode ser empregado para testar hipóteses sobre parâmetros de covariância ou parâmetros de efeito fixo, no contexto de modelos lineares mistos. A estatística é calculada da seguinte forma:

$$2 \log \left(\frac{L_2}{L_1} \right) = 2[\log(L_2) - \log(L_1)], \quad (20)$$

em que L_1 representa a função de verossimilhança do modelo aninhado ou restrito e L_2 , a função de verossimilhança do modelo de referência.

A teoria de verossimilhança determina que sob certas condições de regularidade, a estatística teste da razão de verossimilhanças segue, assintoticamente, a distribuição χ^2 , em que o número de graus de liberdade, gl, é obtido subtraindo-se o número de parâmetros do modelo aninhado do número de parâmetros do modelo de referência. E, para os casos em que os parâmetros encontram-se na fronteira do espaço paramétrico, a estatística teste da razão de verossimilhanças segue uma mistura de χ^2 (SELF; LIANG, 1987).

2.10.1.1 Inferência para parâmetros de efeitos fixos

De acordo com Pinheiro e Bates (2000) e Verbeke e Molenberghs (2000), o teste de uma hipótese linear, sobre os parâmetros de efeito fixo, está baseado na estimação de ML.

Fazendo uso do teste da razão de verossimilhanças, os modelos, aninhado e de referência, devem conter os mesmos componentes de variância e mesmas estruturas para

as matrizes $\mathbf{D}$ e $\mathbf{\Sigma}$, porém diferentes conjuntos de parâmetros fixos. Essa prática permite verificar a importância dos termos fixos do modelo, uma vez que a diferença entre tais modelos encontra-se apenas com relação aos termos fixos presentes.

Neste caso a estatística para o teste da razão de verossimilhanças segue, assintoticamente, a distribuição χ^2 com número de graus de liberdade igual à diferença entre o número de parâmetros de efeito fixo dos modelos em questão.

2.10.1.2 Inferência para os parâmetros de efeitos aleatórios

De acordo com Pinheiro e Bates (2000), os modelos de referência e aninhado, devem ser estimados utilizando o mesmo procedimento. Conforme descrito na Seção 2.6.3, a estimação pelo método REML tende a ser menos viesada do que a estimação pelo método ML e, para os casos de dados balanceados, as estimativas REML coincidem com as obtidas pelo método ANOVA. Por esse último fato, os autores West, Welsh e Galecki (2007) sugerem a estimação dos componentes de variância pelo método REML. Quando o teste é realizado para componentes aleatórios, a especificação da parte fixa deve ser a mesma para os dois modelos.

West, Welsh e Galecki (2007) discriminam dois casos para a realização do teste da razão de verossimilhanças, baseado na verossimilhança restrita. A distribuição da estatística para a hipótese nula depende se os valores dos componentes de variância envolvidos na hipótese estão, ou não, na fronteira do espaço paramétrico.

- (i) **Os parâmetros de covariância referentes à hipótese de nulidade não estão na fronteira do espaço paramétrico:** o interesse está na verificação da homogeneidade de variâncias, ou ainda, se a covariância entre dois efeitos aleatórios é igual a zero. Nesses casos, a estatística segue assintoticamente a distribuição χ^2 com número de graus de liberdade igual à diferença entre o número de parâmetros nos modelos de referência e aninhado.
- (ii) **Os parâmetros de covariância estão na fronteira do espaço paramétrico:** são os casos em que se deseja verificar se um efeito aleatório deve, ou não, permanecer no modelo. Self e Liang (1987), Stram e Lee (1994), Verbeke e Molenbergs (2000) e Butler et. al (2009), afirmam que a estatística para o teste da razão de verossimilhanças para um único parâmetro de variância, o qual se encontra na fronteira do espaço paramétrico, segue uma mistura de distribuições χ^2 , $0, 5\chi_0^2 + 0, 5\chi_1^2$. Para os testes de

razão de verossimilhanças para os casos em que k parâmetros se encontram na fronteira o espaço paramétrico, a estatística segue também uma mistura de distribuições χ^2 , porém, nesse caso a mistura é dada por $0,5\chi_0^2 + 0,5\chi_k^2$.

Self e Liang (1987) apresentaram as distribuições para outros casos, como teste simultâneo para parâmetros de variância para os quais um se encontra na fronteira do espaço paramétrico e outro não.

2.10.2 Testes alternativos para parâmetros de efeito fixo

Verbeke e Molenberghs (2000) e West, Welch e Galecki (2007) estudando os modelos lineares mistos, apresentaram os testes adicionais de Wald, t e Wald- F , que podem ser utilizados para verificar a significância dos termos fixos, ou de uma combinação linear dos mesmos.

2.10.2.1 Teste de Wald

Um teste de Wald aproximado pode ser realizado para cada β_j em $\boldsymbol{\beta}$, com $j = 1, \dots, p$, assim como a obtenção do respectivo intervalo de confiança. Para uma matriz $\mathbf{L}$ conhecida, um teste para as hipóteses

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad H_a : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{0}, \quad (21)$$

considera que

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{L}' \left[\mathbf{L} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1}(\hat{\boldsymbol{\kappa}}) \mathbf{X}_i \right)^{-1} \mathbf{L}' \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

segue, assintoticamente, uma distribuição χ^2 com número de graus de liberdade dados pelo posto da matriz $\mathbf{L}$.

Como o teste de Wald não considera a variabilidade introduzida pela estimação dos componentes de variância ($\hat{\boldsymbol{\kappa}}$), Verbeke e Molenberghs (2000) sugerem a utilização dos testes t e Wald- F para testar hipóteses sobre os parâmetros de efeito fixo.

2.10.2.2 Teste t

O teste t é freqüentemente utilizado para testar hipóteses do tipo:

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_a : \beta_j \neq 0,$$

em que a estatística t é calculada da seguinte forma:

$$t = \frac{\hat{\beta}_j}{\sqrt{\text{Var}(\hat{\beta})}}.$$

No contexto dos modelos lineares mistos, tal estatística não segue uma distribuição t exata e o número de graus de liberdade associados devem ser calculados com base no conjunto de dados, utilizando métodos como o de Satterthwaite (1946), Fai e Cornelius (1996), Kenward e Roger (1997).

2.10.2.3 Teste Wald- F

As hipóteses a serem testadas são conforme 21 e a estatística é definida por:

$$F = \frac{(\hat{\beta} - \beta)' \mathbf{L}' \left[\mathbf{L} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} (\hat{\alpha}) \mathbf{X}_i \right)^{-1} \mathbf{L}' \right]^{-1} \mathbf{L} (\hat{\beta} - \beta)}{\text{posto}(\mathbf{L})}$$

que segue uma distribuição F aproximada, com número de graus de liberdade do numerador dado pelo posto da matriz $\mathbf{L}$ e número de graus de liberdade do denominador calculados utilizando métodos de aproximação como os propostos por Satterthwaite (1946), Fai e Cornelius (1996), Kenward e Roger (1997).

2.10.3 Critérios de informação

Outras estatísticas usadas como medida de comparação da qualidade de ajuste dos modelos são os critérios de informação. Os principais critérios são: Critério de Informação de Akaike (AIC) (AKAIKE, 1974) e Critério de Informação Bayesiano (BIC) (SCHWARZ, 1978). Esses critérios são baseados no logaritmo da função de verossimilhança (ML ou REML) e dependem do número de observações e de parâmetros do modelo.

O AIC consiste em maximizar a função de verossimilhança separadamente para cada modelo, obtendo $l(\hat{\beta}, \hat{\kappa})$, e então calcular valor de AIC, dado por:

$$\text{AIC} = -2l(\hat{\beta}, \hat{\kappa}) + 2d,$$

em que d representa o número total de parâmetros de efeitos fixo e aleatório.

O BIC é um critério semelhante ao AIC, porém penaliza os modelos com maior número de parâmetros. Seu valor é calculado da seguinte forma:

$$\text{BIC} = -2l(\hat{\beta}, \hat{\kappa}) + d \log(n),$$

em que n é o número de observações usadas na estimação do modelo e d , como definido anteriormente.

Um valor baixo para AIC é considerado como representativo de um melhor ajuste e os modelos são selecionados visando obter um mínimo AIC. De forma semelhante interpreta-se BIC.

2.10.4 Estratégias de seleção de modelos

A obtenção de um modelo linear misto adequado passa por um processo de construção, que exige o ajuste de uma série de modelos, com estruturas diferentes, tanto para a parte fixa, quanto para a parte aleatória, ou seja, diferentes estruturas para média e variância.

West, Welch e Galecki (2007) apresentam duas estratégias para seleção do melhor modelo, a estratégia *top-down* e a estratégia *step-up*.

2.10.4.1 Estratégia *top-down*

Esta estratégia, sugerida por Verbeke e Molenberghs (2000) consiste em, inicialmente, extrair toda variação sistemática do modelo. Para tanto, recomenda-se o ajuste do modelo maximal, ou seja, incluir todos os termos de efeitos fixos possíveis e de interesse.

Em seguida, devem-se incluir os termos aleatórios, em que a importância de um termo, ou de uma combinação linear de termos, pode ser verificada fazendo uso do teste da razão de verossimilhanças (estimadas pelo método REML). Note que, nessa fase busca-se a melhor estrutura para a matriz de variâncias e covariâncias $\mathbf{D}$.

Uma vez ajustado o modelo maximal para o componente sistemático e selecionados os termos aleatórios, a variabilidade restante se deve aos resíduos. Logo, seleciona-se uma estrutura apropriada para a matriz de variâncias e covariâncias $\mathbf{\Sigma}$.

O próximo passo consiste em reduzir o modelo maximal, verificando a importância dos termos fixos, bem como dos termos aleatórios, utilizando testes adequados.

2.10.4.2 Estratégia *step-up*

A presente estratégia utilizada por Snijders e Bosker(1999) e por Raudenbush e Bryk (2002), ao contrário da *top-down*, sugere o ajuste da parte fixa do modelo, contendo somente o efeito da média geral, bem como da parte aleatória contendo efeitos dos termos não aleatorizados. Em seguida, são incluídos efeitos de covariáveis como de efeitos fixos e,

possivelmente, efeitos aleatórios associados aos mesmos.

A necessidade da inclusão de termos no modelo pode ser verificada pelo teste da razão de verossimilhanças ou de critérios de informação.

2.11 Diagnósticos

Diagnósticos devem ser parte do processo de construção de modelos e análise de conjuntos de dados. Técnicas informais são, comumente, utilizadas para diagnósticos de resíduos, como a construção de gráficos para verificar pressuposições do modelo, detectar *outliers* e possíveis observações influentes (WEST; WELCH; GALECKI, 2007).

2.11.1 Resíduos

Os resíduos são utilizados para examinar as suposições do modelo e detectar a presença de *outliers* e observações influentes. No contexto dos modelos lineares mistos pode-se distinguir o resíduos marginal ($\mathbf{r}_m$) e condicional ($\mathbf{r}_c$), apresentados a seguir:

(i) Resíduo marginal:

$$\mathbf{r}_m = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}, \quad (22)$$

que consiste da diferença entre valor observado e a média marginal estimada.

(ii) Resíduo condicional:

$$\mathbf{r}_c = \hat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}, \quad (23)$$

consiste da diferença entre valor observado e valor predito da observação.

Os resíduos $\mathbf{r}_m$ e $\mathbf{r}_c$ não são adequados para diagnósticos, pois podem apresentar correlações, mesmo para dados não correlacionados, e possuem difícil interpretação quando são incorporadas variâncias distintas ao modelo. Um modo de minimizar tais problemas consiste em trabalhar com os resíduos padronizados, ou seja, dividir os resíduos pelos verdadeiros valores referentes aos desvios padrões (para $\mathbf{r}_m$ e $\mathbf{r}_c$). Entretanto, tais valores, geralmente, são desconhecidos, e os resíduos são divididos pela estimativa dos desvios padrões, obtendo-se assim os resíduos estudentizados, descritos a seguir:

(iii) Resíduo marginal estudentizado:

$$\mathbf{r}_m^{estudentizado} = \frac{\mathbf{r}_m}{\sqrt{\widehat{\text{Var}}[\mathbf{r}_m]}}. \quad (24)$$

(iv) Resíduo condicional estudentizado:

$$\mathbf{r}_c^{estudentizado} = \frac{\mathbf{r}_c}{\sqrt{\widehat{\text{Var}}[\mathbf{r}_c]}}. \quad (25)$$

2.11.2 Distribuição dos resíduos e dos efeitos aleatórios

Uma escolha natural para diagnosticar os efeitos aleatórios é considerar o preditor descrito na equação 11, que é o melhor preditor linear não viesado de $\mathbf{b}$, fazendo uso de gráficos de diagnósticos padrões, ou seja, histogramas, gráficos de quantil-quantil e gráficos de dispersão. Mas estudar a normalidade desses preditores fornece um diagnóstico limitado, pois suas distribuições não, necessariamente, refletem a verdadeira distribuição dos efeitos aleatórios.

Uma alternativa proposta por Lange e Ryan (1989) consistiu da utilização de gráficos quantil-quantil não para o próprio EBLUP (*empirical best linear unbiased predictor* - melhor preditor linear não viesado empírico), mas para uma padronização desses, na qual $\hat{\mathbf{b}}$ é ponderado pela raiz quadrada de sua variância estimada. Essa técnica recebeu críticas de vários autores, entre eles Verbeke e Molenberghs (2000), que afirmam que uma vez que essa padronização dependa de $\hat{\mathbf{b}}$ e de $\hat{\epsilon}$, tal gráfico serve apenas para indicar que a padronização não segue a distribuição esperada sob o modelo assumido.

Assim como para os efeitos aleatórios, West, Welch e Galecki (2007) sugerem a utilização de gráficos de quantil-quantil e gráficos de dispersão para a verificação da normalidade dos resíduos, e nesse caso, dos resíduos condicionais estudentizados. De acordo com Hilden-Minton (1995) e Nobre (2004), quando o interesse está em verificar a suposição de normalidade dos resíduos, a utilização dos resíduos condicionais não é aconselhável, pois ele é confundido com $\mathbf{b}$, logo quando $\mathbf{b}$ se afasta muito da normalidade, os resíduos condicionais podem não apresentar características de normalidade, mesmo quando ϵ segue uma distribuição normal. Hilden-Minton (1995) e Nobre (2004) sugerem a utilização dos resíduos denominados de confundimento mínimo e a análise visual de gráficos de quantil-quantil para os mesmos.

Quando considerada a abordagem por modelos mistos, testes clássicos de normalidade podem não ser adequados, pois os erros não são independentes. Hwang e Wei (2006) propuseram um método de transformação para converter dados correlacionados em dados não correlacionados, considerando um experimento contendo dois estágios de agrupamento. Como resultado, testes clássicos de normalidade podem ser utilizados.

Claeskens e Hart (2009) fornecem várias estratégias para construção de testes estatísticos da hipótese de normalidade dos efeitos aleatórios e/ou distribuição dos erros. Esses testes são não paramétricos, pois os autores não assumem uma única forma paramétrica para o modelo alternativo.

2.11.3 Resíduos com confundimento mínimo

Hilden-Minton (1995) propôs uma abordagem para detecção de observações influentes e verificação das suposições do modelo, considerando modelos multiníveis.

De acordo com o mesmo autor, existem três tipos de distúrbios (perturbações) no modelo, o erro marginal, o erro condicional e os efeitos aleatórios. Resíduos que em suas funções dependem de algum distúrbio adicional, são denominados resíduos confundidos, que é o caso, por exemplo, do resíduo condicional, que está confundido com o vetor de parâmetros de efeito aleatório.

O autor sugeriu trabalhar com uma combinação linear dos resíduos condicionais, $\mathbf{l}'\mathbf{r}_c$, de modo a minimizar o confundimento entre $\boldsymbol{\epsilon}$ e $\mathbf{b}$. E então, realizar análises visuais, utilizando gráficos como de dispersão e quantil-quantil para tais resíduos de confundimento mínimo.

Para a obtenção de $\mathbf{l}'\mathbf{r}_c$, sejam as seguintes matrizes:

$$\begin{aligned}\boldsymbol{\Sigma} &= \sigma^2 \mathbf{R} \\ \mathbf{D} &= \sigma^2 \mathbf{G} \\ \mathbf{M} &= (\mathbf{ZGZ}' + \mathbf{R})^{-1} \\ \mathbf{Q} &= \mathbf{M} - \mathbf{MX}(\mathbf{X}'\mathbf{MX})^{-1}\mathbf{X}'\mathbf{M},\end{aligned}$$

seja também $\mathbf{d} = \mathbf{Z}\hat{\mathbf{b}}$ e $\boldsymbol{\zeta} = \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$, tal que a matriz de variâncias e covariâncias de $\boldsymbol{\zeta}$ é $\mathbf{V} = \sigma^2(\mathbf{ZGZ}' + \mathbf{R})$.

Desse modo $\mathbf{r}_m = \mathbf{r}_c + \mathbf{d}$, e observe que:

$$\begin{aligned}\mathbf{r}_c &= \mathbf{RQy} \\ \mathbf{r}_m &= \mathbf{M}^{-1}\mathbf{Qy} \\ \mathbf{d} &= \mathbf{ZGZ}'\mathbf{Qy}.\end{aligned}$$

Com as respectivas variâncias dadas por:

$$\begin{aligned}\text{Var}(\mathbf{r}_c) &= \sigma^2 \mathbf{RQM}^{-1}\mathbf{QR} \\ &= \sigma^2 \mathbf{RQR},\end{aligned}$$

$$\begin{aligned}\text{Var}(\mathbf{r}_m) &= \sigma^2 \mathbf{M}^{-1}\mathbf{QM}^{-1} \\ &= \sigma^2 [\mathbf{ZGZ}' + \mathbf{R} - \mathbf{X}(\mathbf{X}'\mathbf{MX})^{-1}\mathbf{X}'],\end{aligned}$$

$$\begin{aligned}\text{Var}(\mathbf{d}) &= \sigma^2 \mathbf{ZGZ}'\mathbf{QM}^{-1}\mathbf{QZGZ}' \\ &= \sigma^2 \mathbf{ZGZ}'\mathbf{QZGZ}'.\end{aligned}$$

Logo,

$$\mathbf{r}_c - \boldsymbol{\epsilon} = \mathbf{RQZb} + (\mathbf{RQ} - \mathbf{I})\boldsymbol{\epsilon} \quad (26)$$

$$\mathbf{d} - \mathbf{Zb} = (\mathbf{ZGZ}'\mathbf{Q} - \mathbf{I})\mathbf{Zb} + \mathbf{ZGZ}'\mathbf{Q}\boldsymbol{\epsilon}. \quad (27)$$

Da equação 26 nota-se que $\mathbf{r}_c$ está confundido com $\mathbf{b}$. De modo análogo, da equação 27, pode-se concluir que $\mathbf{d}$ está confundido com $\boldsymbol{\epsilon}$.

A habilidade de verificar a não normalidade em $\boldsymbol{\epsilon}$ a partir do exame de $\mathbf{r}_c$ diminui a medida que a variância de $\mathbf{RQZb}$ aumenta em relação à variância de $\mathbf{RQ}\boldsymbol{\epsilon}$, uma vez que $\mathbf{r}_c = \mathbf{RQ}\boldsymbol{\epsilon} + \mathbf{RQZb}$ (HILDEN-MINTON, 1995). Definiu-se então, a fração de confundimento ($CF_{\mathbf{r}_c}$), dada por:

$$CF_{\mathbf{r}_c} = \frac{\mathbf{u}'_i \text{Var}(\mathbf{RQZb}) \mathbf{u}_i}{\mathbf{u}_i \text{Var}(\mathbf{r}_c) \mathbf{u}_i} = \frac{\mathbf{u}'_i \mathbf{RQZGZ}'\mathbf{QRu}_i}{\mathbf{u}_i \mathbf{RQRu}_i} = 1 - \frac{\mathbf{u}'_i \mathbf{RQRQRu}_i}{\mathbf{u}_i \mathbf{RQRu}_i},$$

em que, $\mathbf{u}_i$ é a i -ésima coluna de uma matriz identidade de ordem n .

Como o interesse foi obter uma fração de confundimento mínimo, ou seja, uma transformação linear, $\mathbf{l}'\mathbf{r}_c$, dos resíduos condicionais, tal que essa fração seja mínima e consequentemente, que a razão,

$$\lambda_i = \frac{\mathbf{l}'_i \mathbf{RQRQRl}_i}{\mathbf{l}'_i \mathbf{RQRl}_i}, \quad (28)$$

assuma seu valor máximo, esta solução remete a um problema de autovalores de matrizes positivas semi-definidas.

Sabe-se que $0 \leq \lambda_i \leq 1$, logo o vetor de transformação linear $\mathbf{l}_i$ será restrito, tal que,

$$\mathbf{l}_i' \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{l}_i > 0. \quad (29)$$

Se $\mathbf{R}$ é uma matriz de posto completo, segundo a restrição 29, o foco é o espaço não nulo da matriz positiva semi-definida $\mathbf{Q}$. E, de acordo com as propriedades dos autovalores e autovetores da matriz $\mathbf{R} \mathbf{Q}$ apresentadas por Hilden-Minton (1995), e pela decomposição em valores singulares (HARVILLE, 1997), pode-se obter:

$$\mathbf{R}^{1/2} \mathbf{Q} \mathbf{R}^{1/2} = \mathbf{K} \mathbf{\Lambda} \mathbf{K}' \quad \text{e} \quad \mathbf{K}' \mathbf{K} = \mathbf{I},$$

em que, $\mathbf{K}$ é uma matriz de ordem $n \times (n - p_X)$, $\mathbf{\Lambda}$ é uma matriz quadrada de ordem $(n - p_X)$, $(n - p_X)$ é o número de autovalores não nulos de $\mathbf{R}^{1/2} \mathbf{Q} \mathbf{R}^{1/2}$ e p_X é o posto da matriz $\mathbf{X}$ do modelo 3.

Seja,

$$\mathbf{l}_i = \mathbf{R}^{-1/2} \mathbf{K} \mathbf{\Lambda}^{-1/2} \mathbf{v}_i, \quad (30)$$

substituindo-se 30 em 28, para algum $\mathbf{v}_i$ de ordem $(n - p_X) \times 1$, λ_i pode ser reescrito como:

$$\lambda_i = \frac{\mathbf{v}_i' \mathbf{\Lambda} \mathbf{v}_i}{\mathbf{v}_i' \mathbf{v}_i}.$$

Assim, $\lambda_{n-p_X} \leq \lambda_i \leq \lambda_1$, com $\lambda_{n-p_X} \leq \dots \leq \lambda_1 \leq 1$, representando os autovalores ordenados de $\mathbf{R}^{1/2} \mathbf{Q} \mathbf{R}^{1/2}$.

Seja $\mathbf{v}_i$ a i -ésima coluna de uma matriz identidade de ordem $(n - p_X)$, logo,

$$\mathbf{l}_i = \mathbf{R}^{-1/2} \mathbf{K} \mathbf{\Lambda}^{-1/2} \mathbf{v}_i = \lambda_i^{-1/2} \mathbf{R}^{-1/2} \mathbf{K}_i \quad i = 1, \dots, n - p_X,$$

em que $\mathbf{K}_i$ representa a i -ésima coluna de $\mathbf{K}$.

Portanto,

$$\begin{aligned} \mathbf{l}_i' \mathbf{r}_c &= (\lambda_i^{-1/2} \mathbf{R}^{-1/2} \mathbf{K}_i)' \mathbf{R} \mathbf{Q} \mathbf{y} \\ &= \lambda_i^{-1/2} \mathbf{K}_i' \mathbf{R}^{1/2} \mathbf{Q} \mathbf{y}. \end{aligned}$$

Com,

$$\begin{aligned} \text{Cov}(\mathbf{l}_i' \mathbf{r}_c, \mathbf{l}_j' \mathbf{r}_c) &= \sigma^2 \mathbf{l}_i' \mathbf{R} \mathbf{Q} \mathbf{r} \mathbf{l}_j \\ &= \begin{cases} \sigma^2 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases} \end{aligned}$$

Logo, $\frac{\mathbf{l}'_i \mathbf{r}_c}{\sigma}$, $i = 1, \dots, (n - p_X)$, são os resíduos não correlacionados, com fração de confundimento $1 - \lambda_i$, denominados por Hilden-Minton (1995), resíduos com confundimento mínimo, por meio dos quais pode-se verificar a hipótese de normalidade dos resíduos condicionais.

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Experimento com instalações avícolas

Visando o conforto térmico e o bem estar dos animais em instalações avícolas, Castro (2011) realizou uma pesquisa na área experimental do Núcleo de Pesquisa em Ambiência (NUPEA), pertencente ao Departamento de Engenharia de Biossistemas (LEB) da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP), com o objetivo de avaliar a eficiência de forros como elementos atenuadores da radiação solar. Para tanto, foram utilizados dois tipos diferentes de telha (cerâmica - C e fibrocimento - F) associados a dois tipos de forro (A e B).

As quatro combinações entre tipos de telha e tipos de forro foram aleatorizados a quinze blocos, e cada unidade experimental, sendo esta uma instalação de confinamento, teve registros de temperatura (Temperatura de bulbo seco, °C), umidade relativa (%) e temperatura de globo negro (°C) às 8h00min, 11h00min, 14h00min e 17h00min. Nos mesmos horários foi registrada a velocidade do vento, a fim de calcular a carga térmica radiante. Logo, para cada hora foram determinados: Temperatura de bulbo seco (°C), Umidade relativa (%), Temperatura de bulbo negro (°C), Entalpia específica (h, kJ/kg de ar seco), Índice de temperatura de globo e umidade e carga térmica radiante. Neste trabalho, apenas a variável temperatura de bulbo seco (°C) foi analisada.

3.1.2 Experimento com polpas de Eucalipto

Pereira (1969) descreveu um experimento realizado com povoamentos de eucalipto, para investigar diferenças entre as polpas e as características do papel produzido. O material utilizado na experimentação foi obtido de árvores das espécies *E. saligna* SM, de 5 e 7 anos, *E. alba* Reinw, de 5 e 7 anos e *E. grandis* Hill ex Maiden, de 7 anos, retiradas de povoamentos pertencentes à Champion Celulose S. A., em Mogi Guaçu, Estado de São Paulo.

No presente estudo, somente os dados referentes às espécies *E. saligna* e *E. alba* foram analisados, dos quais foram escolhidos os talhões mais representativos no tocante ao desenvolvimento médio das espécies naquelas idades, para aquele local. Três lotes de cinco árvores de cada uma das quatro áreas (combinação entre espécies e idades) foram selecionados.

O experimento original foi realizado em três fases, que são descritas a seguir:

Na **primeira fase** as toras de cinco árvores de cada um dos doze lotes foram dispostas, ao acaso, sobre a esteira rolante que alimentava o picador industrial. Por três vezes o picador foi desligado, amostrando-se quatro grupos dessa mistura de cavacos, após misturar convenientemente.

Na **segunda fase**, os cavacos produzidos para cada combinação entre espécie e idade foram transportados para um laboratório de celulose e papel, onde foram cozidos em um digestor de aço inoxidável. Cada grupo de cavacos foi cozido separadamente, e a partir de cada cozimento, seis amostras foram coletadas.

Na **terceira fase**, foi realizada a refinação da celulose em um moinho Jokro-Muhls. A refinação foi conduzida em seis tempos, a partir de 30 minutos e a intervalos iguais de 30 minutos, até 180 minutos. Cada tempo de refinação foi aleatorizado a cada uma das seis posições no moinho. Cada cozimento foi processado em uma das quarenta e oito rodadas e cada uma das seis amostras foi aleatorizada a uma das seis posições.

Para cada tempo de refinação foram determinados: grau de refinação ($^{\circ}\text{SR}$), resistência à tração (comprimento de auto-ruptura), resistência ao arrebentamento (índice de arrebentamento), resistência ao rasgo (índice de rasgo), resistência ao dobramento (número de dobras duplas) e peso específico aparente (g/cm^3). A variável grau de refinação ($^{\circ}\text{SR}$) foi analisada como motivação do estudo de Brien e Demétrio (2009), e a variável resposta analisada neste trabalho foi peso específico aparente (g/cm^3).

3.2 Métodos

Inicialmente, para que os conjuntos de dados pudessem ser analisados de acordo com a proposta dos modelos lineares mistos (LAIRD; WARE, 1982), foram identificados os fatores presentes em cada experimento, discriminando-os como fator aleatorizado, fator não aleatorizado (BRIEN, 2007) ou fator longitudinal (BRIEN; DEMÉTRIO, 2009). Foram identificadas também suas camadas (BRIEN, 1983) e as respectivas fórmulas estruturais, indicando assim a relação existente entre os mesmos.

Como forma de identificar um possível confundimento entre os fatores em camadas distintas, construiu-se o diagrama de aleatorização (BRIEN; BAILEY, 2006), ainda com o mesmo objetivo, foram construídas tabelas de decomposição do número de graus de liberdade, os quais foram calculados fazendo uso dos diagramas de Hasse (TAYLOR, HILTON, 1981).

Em seguida, utilizou-se a abordagem proposta por Brien e Demétrio (2009) para a construção dos modelos mistos, formulando assim os modelos homogêneos fixo e aleatório e os modelos gerais, fixo e aleatório. Com o objetivo de se detectar tendências e conhecer um pouco a respeito da variabilidade dos dados, foram construídos os gráficos de perfis, considerando-se cada combinação de interesse (tratamento).

O primeiro modelo ajustado, em termos das notações propostas, foi aquele que assume F, para a parte fixa, ou seja, sem considerar tendências sistemáticas, e HR para a parte aleatória, excluindo-se fatores totalmente confundidos com outros já presentes no modelo. Esse primeiro modelo caracteriza o modelo de componentes de variância, denominado também como modelo de ANOVA ou Base.

Optou-se pelo método de estimação da máxima verossimilhança restrita (PETTERSON; THOMPSON, 1971), para o vetor de componentes de variância, e pelo método dos mínimos quadrados generalizados, para a estimação do vetor de parâmetros de efeito fixo.

A estratégia para a seleção de modelos adotada nesse estudo foi a *top-down* (VERBEKE; MOLENBERGHS, 2000), portanto, o segundo modelo ajustado foi aquele composto por GF para a parte fixa, no qual foram incluídas tendências sistemáticas com respeito à variável correspondente ao fator longitudinal. GF pode ser visto como o modelo maximal para os termos de efeito fixo. Com relação a parte aleatória, esse segundo modelo considera GR, em que o termo $uc(\text{Fator Longitudinal})$ pode permitir, a princípio, tendências variando aleatoriamente.

Conforme a estratégia adotada, em um primeiro momento foram selecionados fatores aleatórios e suas estruturas, considerando-se a matriz de variâncias e covariâncias $\mathbf{D}$, conseqüentemente, os elementos que compõem o vetor $\boldsymbol{\gamma}$.

A comparação dos modelos foi realizada da seguinte maneira:

- (i) Para os casos em que o interesse consistiu em verificar se um efeito aleatório deve, ou não, permanecer no modelo, foi calculada a razão de verossimilhanças entre o modelo de referência e o modelo aninhado, e testou-se a estatística em questão com uma mistura de distribuições χ^2 , do tipo $\frac{1}{2}\chi_q^2 + \frac{1}{2}\chi_{q+1}^2$, em que q é a ordem do vetor $\mathbf{b}$ do modelo aninhado.
- (ii) Para os casos em que o interesse consistiu em verificar a homogeneidade de variâncias, foi calculada a razão de verossimilhanças entre os modelos de referência e aninhado, porém a estatística foi testada utilizando-se a distribuição χ^2 , com número de graus de

liberdade iguais à diferença entre o número de parâmetros nos modelos de referência e aninhados.

Uma vez selecionados os componentes e a estrutura da matriz $\mathbf{D}$, de acordo com o método *top-down*, deve-se verificar a estrutura da matriz $\mathbf{\Sigma}$, assim, nesse estágio, foram ajustados modelos que assumem estruturas do tipo NE, $\text{AR}_H(1)$, $\text{AR}(1)$, SC, em bandas, entre outras.

Tendo encontrado estruturas adequadas para as matrizes $\mathbf{D}$ e $\mathbf{\Sigma}$, por meio do teste Wald- F foi verificada a importância de cada um dos termos fixos. Sendo necessário o ajuste de um modelo reduzido, a importância dos termos aleatórios, bem como suas estruturas foram verificadas novamente.

Os critérios de informação, AIC (AKAIKE, 1974) e BIC (SCHWARZ, 1978), auxiliaram na seleção de modelos encaixados e foram decisivos para a seleção de modelos não encaixados.

Diagnósticos foram feitos observando-se os gráficos de dispersão para os resíduos condicionais estudentizados e para o EBLUP. A verificação da normalidade tanto para o vetor de parâmetros de efeito aleatório quanto para os resíduos, foi feita utilizando os gráficos de quantil-quantil com envelope de simulação, porém os mesmos foram construídos para os resíduos com confundimento mínimo (HILDEN-MINTON, 1995) e para o EBLUP com confundimento mínimo, este último descrito conforme a subseção a seguir.

3.3 EBLUP com confundimento mínimo

Conforme desenvolvimento do resíduo com confundimento mínimo, este trabalho tem como proposta a utilização do EBLUP com confundimento mínimo, com os quais pode-se construir gráficos de quantil-quantil com envelope de simulação, e realizar análises visuais.

Da equação 27, tem-se:

$$\mathbf{d} = \mathbf{ZGZ}'\mathbf{QZb} + \mathbf{ZGZ}'\mathbf{Q}\epsilon,$$

logo, pretende-se quantificar a variabilidade de $\mathbf{d}$ devido à influência de ϵ . Assim, a fração de confundimento é dada por:

$$CF_{\mathbf{d}_i} = \frac{\mathbf{u}'_i \text{Var}(\mathbf{ZGZ}'\mathbf{Q}\epsilon) \mathbf{u}_i}{\mathbf{u}'_i \text{Var}(\mathbf{d}) \mathbf{u}_i} = \frac{\mathbf{u}'_i \mathbf{ZGZ}'\mathbf{Q}\mathbf{RQZGZ}' \mathbf{u}_i}{\mathbf{u}'_i \mathbf{ZGZ}'\mathbf{QZGZ}' \mathbf{u}_i} = 1 - \frac{\mathbf{u}'_i \mathbf{ZGZ}'\mathbf{QZGZ}'\mathbf{QZGZ}' \mathbf{u}_i}{\mathbf{u}'_i \mathbf{ZGZ}'\mathbf{QZGZ}' \mathbf{u}_i}$$

em que, $\mathbf{u}_i$ representa a i -ésima coluna de uma matriz identidade de ordem n .

Neste caso, deseja-se obter uma fração de confundimento mínimo, isto é, uma transformação linear, $\mathbf{l}'_{2i}\mathbf{d}$, tal que a razão dada por:

$$\lambda_{2i} = \frac{\mathbf{l}'_{2i}\mathbf{ZGZ}'\mathbf{QZGZ}'\mathbf{QZGZ}'\mathbf{l}_{2i}}{\mathbf{l}'_{2i}\mathbf{ZGZ}'\mathbf{QZGZ}'\mathbf{l}_{2i}} \quad (31)$$

assuma seu valor máximo.

Sabe-se que, $0 \leq \lambda_{2i} \leq 1$, desse modo, a transformação linear será restrita a

$$\mathbf{l}'_{2i}\mathbf{ZGZ}'\mathbf{QZGZ}'\mathbf{l}_{2i} > 0. \quad (32)$$

Com foco no espaço não nulo de $\mathbf{Q}$, sujeito a restrição 32 e, de modo análogo à obtenção dos resíduos com confundimento mínimo, tem-se:

$$(\mathbf{ZGZ}')^{1/2}\mathbf{Q}(\mathbf{ZGZ}')^{1/2} = \mathbf{K}_2\mathbf{\Lambda}_2\mathbf{K}'_2 \quad \text{e} \quad \mathbf{K}'_2\mathbf{K}_2 = \mathbf{I},$$

em que $\mathbf{K}_2$ é uma matriz de ordem $n \times (m - p_X)$ e $\mathbf{\Lambda}_2$ é uma matriz quadrada de ordem $(m - p_X)$, $m = \text{posto}[\mathbf{X}|\mathbf{Z}]$ e $p_X = \text{posto}(\mathbf{X})$.

Seja,

$$\mathbf{l}_{2i} = (\mathbf{ZGZ}')^{-1/2}\mathbf{K}_2\mathbf{\Lambda}_2^{-1/2}\mathbf{v}_i, \quad (33)$$

substituindo-se 33 em 31, para algum $\mathbf{v}_i$ de ordem $(m - p_X) \times 1$, λ_{2i} pode ser reescrito como

$$\lambda_{2i} = \frac{\mathbf{v}'_i\mathbf{\Lambda}_2\mathbf{v}_i}{\mathbf{v}'_i\mathbf{v}_i}.$$

Assim, $\lambda_{2_{m-p_X}} \leq \lambda_{2i} \leq \lambda_{2_1}$ com $\lambda_{2_{m-p_X}} \leq \dots \leq \lambda_{2_1} \leq 1$ representando os autovalores ordenados de $(\mathbf{ZGZ}')^{1/2}\mathbf{Q}(\mathbf{ZGZ}')^{1/2}$.

Seja $\mathbf{v}_i$ a i -ésima coluna de uma matriz identidade de ordem $(m - p_X)$, então:

$$\mathbf{l}_{2i} = (\mathbf{ZGZ}')^{-1/2}\mathbf{K}_2\mathbf{\Lambda}_2^{-1/2}\mathbf{v}_i = \lambda_{2i}^{-1/2}(\mathbf{ZGZ}')^{-1/2}\mathbf{K}_{2i}, \quad i = 1, \dots, (m - p_X),$$

em que $\mathbf{K}_{2i}$ representa a i -ésima coluna de $\mathbf{K}_2$.

Portanto,

$$\mathbf{l}'_{2i}\mathbf{d} = \lambda_{2i}^{-1/2}\mathbf{K}'_{2i}(\mathbf{ZGZ}')^{1/2}\mathbf{Qy}$$

e

$$\begin{aligned} \text{Cov}(\mathbf{l}'_{2i}\mathbf{d}, \mathbf{l}'_{2j}\mathbf{d}) &= \mathbf{l}'_{2i}\mathbf{ZGZ}'\mathbf{QZGZ}'\mathbf{l}_{2j} \\ &= \begin{cases} \sigma^2 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}. \end{aligned}$$

Logo, $\frac{\mathbf{l}'_{2i}\mathbf{d}}{\sigma}$, $i = 1, \dots, (m - p_X)$, são os $\mathbf{Z}\hat{\mathbf{b}}$ não correlacionados, com fração de confundimento $1 - \lambda_{2i}$, a serem denominados por $\mathbf{Z}\hat{\mathbf{b}}$ com confundimento mínimo (ou EBLUP com confundimento mínimo), por meio dos quais pode-se verificar a normalidade de $\mathbf{b}$.

3.4 Pacote estatístico

Para a realização das análises e construção dos gráficos mencionados utilizou-se o pacote ASReml-R (BUTLER; CULLIS; GILMOUR, 2009) e as funções descritas na Seção 3.4.2.

3.4.1 ASReml

ASReml é um poderoso *software* estatístico em constante desenvolvimento, e é um trabalho conjunto entre *Biometrics Program of NSW Department of Primary Industries* e *Biomathematics and Bioinformatics Division of Rothamsted Research*.

Construído para ajustar modelos lineares mistos para grandes conjuntos de dados utilizando o método da máxima verossimilhança restrita, o ASReml já está em sua terceira versão, e permite ao usuário o ajuste das mais diversas e complexas estruturas para as matrizes de variâncias e covariâncias $\mathbf{D}$ e $\mathbf{\Sigma}$, bem como a inclusão, nos casos de experimentos de melhoramento genético, seja este animal ou vegetal, de matrizes de parentesco.

Disponível também em sua terceira versão, a biblioteca `asreml` (ASReml-R) é uma interface para o *software* R, que apresenta o conveniente uso do R para manipulação de dados e produção de gráficos combinado com o poder do ASReml. Tal biblioteca é uma alternativa ao uso das bibliotecas `lme` (PINHEIRO; BATES, 2000) e `lme4` (BATES; MAECHLER; BOLKER, 2011).

A função `asreml` é composta pelos seguintes argumentos:

```
asreml(fixed = y ~ 1, random, sparse, rcov, G.param, R.param,
       predict = predict.asreml(), constraints = asreml.constraints(),
       data = sys.parent(), subset, family = asreml.gaussian(),
       weights = NULL, offset = NULL, na.method.Y = "include",
       na.method.X = "fail", keep.order = FALSE, ran.order = "user",
       fixgammas = FALSE, as.multivariate = NULL, model.frame = FALSE,
       start.values = FALSE, dump.model = FALSE, model = FALSE,
       control = asreml.control(...), ...)
```

em que:

- `fixed` especifica a fórmula para os termos de efeito fixo do modelo, de acordo com a construção dos modelos mistos, apresentada na Seção 2.9, `fixed =  $\mathbf{y} \sim \text{GF}$` ;

- **random** especifica a fórmula para os termos de efeito aleatório, exceto o termo para o indivíduo (unidade de observação). Nesse argumento serão trabalhadas as estruturas para a matriz $\mathbf{D}$. De acordo com as notações da Seção 2.9, **random** = $\sim$ GR;
- **rcov** especifica o termo para o indivíduo e a estrutura para a matriz de variâncias e covariâncias Σ ;

Cabe salientar que os operadores de tendências e de correlações desiguais, apresentadas na Tabela 5, entre outros, são funções já programadas na biblioteca **asreml**. Maiores informações podem ser obtidas junto a documentação da biblioteca.

Seus principais objetos são:

- BLUP do vetor de parâmetros de efeito aleatório;
- estimativas de mínimos quadrados generalizados do vetor de parâmetros de efeito fixo;
- estimativas de máxima verossimilhança restrita dos componentes de variância;
- predições a partir do modelo linear, se solicitado;
- caso solicitado, por meio do argumento **aom=T**, também irão compor os objetos, os resíduos condicionais estudentizados e o BLUP estudentizado.

Uma particularidade, tanto do *software* ASReml quanto da biblioteca **asreml**, é que existe um custo para se obter suas licenças, entretanto uma versão *discovery* está disponível em <http://www.vsnr.co.uk/software/discovery/asreml-discovery>.

3.4.2 Novas funções

Embora a função **asreml** apresente em seus objetos $\hat{\mathbf{b}}$, $\hat{\beta}$ e as estimativas dos componentes de variância, $\hat{\mathbf{k}}$, a função não apresenta as matrizes $\mathbf{X}$, $\mathbf{Z}$, $\hat{\mathbf{D}}$ ou $\hat{\Sigma}$, para que possam ser calculados os resíduos com confundimento mínimo e o EBLUP com confundimento mínimo. Para tanto, foram criadas funções auxiliares, a serem compiladas no *software* R, utilizando os objetos disponíveis da função **asreml**, de tal modo a se obter as matrizes de delineamento, $\mathbf{X}$ e $\mathbf{Z}$ e as matrizes de variâncias e covariâncias $\hat{\mathbf{D}}$ e $\hat{\Sigma}$.

Para criar as quatro matrizes de interesse, devido ao tempo computacional, foram desenvolvidas duas funções, a primeira função, **matr**, disponibiliza as matrizes $\mathbf{X}$, $\mathbf{Z}$ e $\hat{\mathbf{D}}$, enquanto que a função **matrR** disponibiliza a matriz $\hat{\Sigma}$. Seus argumentos são:

```
matr<- function(modelo, dados, long, Long){...      }
```

e

```
matrR<- function(modelo, dados, long, Long){...      }
```

em que:

- **modelo** indica o modelo ajustado utilizando a função **asreml**, com algumas particularidades. Todo modelo deve conter o argumento **rcov= ~ termo indivíduo** e o argumento **aom=T**;
- **dados** indica os dados utilizados para o ajuste do modelo, inseridos no R por meio da função **asreml.read.table(...)**. Uma informação importante é que dados lidos a partir desta função possuem suas colunas classificadas como fatores qualitativos se seu nome possuir letra inicial maiúscula e, colunas classificadas como fatores quantitativos se seu nome possuir letra inicial minúscula. Qualquer outro fator incluído no conjunto de dados após a leitura no R deve ser classificado como fator qualitativo, caso seja o interesse;
- **long** indica o fator longitudinal quantitativo, como por exemplo **long=c('tempo')**, e
- **Long** indica o fator longitudinal como qualitativo, como por exemplo **Long=c('Tempo')**.
- Os termos **long** e **Long** devem ser criados e estarem disponíveis no conjunto de dados.

As funções **matr** e **matrR** foram programadas para reconhecer todos os operadores de tendências e de correlações desiguais, apresentados na Tabela 5.

A função **lc** tem como objetos, os resíduos com confundimento mínimo (**Le**) e o EBLUP com confundimento mínimo (**Ld**), e seus argumentos são:

```
lc<- function(modelo, dados, long, Long, y, ue) {...      }
```

em que:

- **modelo, dados, long, Long**: conforme funções **matr** e **matrR**;
- **y** é o vetor da variável resposta;
- **ue** é o vetor, de mesma ordem de **y**, identificando a unidade experimental.

A função **graficos** tem como objetos os seguintes gráficos:

- índices *versus* EBLUP;
- fator longitudinal *versus* resíduo marginal;
- unidade experimental *versus* resíduo condicional estudentizado;
- resíduo com confundimento mínimo com envelope de simulação;
- EBLUP com confundimento mínimo com envelope de simulação;

E seus argumentos são:

```
graficos<- function(modelo, dados, long, Long, y, ue) {...      }
```

ou seja, os mesmos argumentos da função `lc`.

As funções implementadas neste trabalho podem ser solicitadas via e-mail: alcarde83@yahoo.com.br.

4 RESULTADOS E DISCUSSÃO

4.1 Experimento com instalações avícolas

O experimento com instalações avícolas é composto por apenas duas camadas, a primeira que contém os fatores aleatorizados {Telhas, Forros}, com a seguinte fórmula estrutural intra-camada:

$$(2 \text{ Telhas} * 2 \text{ Forros}). \quad (34)$$

A segunda camada contém os fatores {Blocos, Parcelas, Horas}, tal que:

$$(15 \text{ Blocos} / 4 \text{ Parcelas}) * 4 \text{ Horas}, \quad (35)$$

com Horas sendo o único fator longitudinal presente no experimento e Parcelas (dentro de Blocos) as unidades experimentais.

As letras iniciais dos fatores foram utilizadas para denotar os mesmos, tal que T(elhas), F(orros), B(locos), P(arcelas), H(oros).

Na Figura 5 pode ser observado o diagrama de aleatorização para o presente experimento, o qual apresenta uma única aleatorização das combinações entre níveis de Telhas e Forros às Parcelas dentro de Blocos.

Figura 5 - Diagrama de aleatorização para o experimento com instalações avícolas

A partir das fórmulas estruturais 34 e 35, têm-se os seguintes modelos intra-camada aleatório e intra-camada fixo,

$$\text{IR : } (\text{Blocos} / \text{Parcelas}) * \text{Horas},$$

$$\text{IF : } \text{Telhas} * \text{Forros}.$$

Sendo as interações Telhas#Horas, Forros#Horas e Telhas#Forros#Horas, de interesse prático, foi considerada a relação cruzada entre os fatores Telhas, Forros e Horas. E, assumindo-se estes fatores como de efeito fixo, e os demais fatores, Blocos e Parcelas, como de efeito aleatório, os modelos HR e F foram descritos por:

$$\text{HR : } \text{Blocos} / \text{Horas} + \text{Blocos} \wedge \text{Parcelas} / \text{Horas},$$

$$\text{F : } \text{Telhas} * \text{Forros} * \text{Horas}.$$

Embora o experimento contenha apenas duas camadas e apresente uma única aleatorização, para uma melhor compreensão dos efeitos confundidos entre os modelos HR e F, construiu-se a tabela de decomposição dos números de graus de liberdade (Tabela 6), os quais foram obtidos com o auxílio dos diagramas de Hasse, apresentados nas Figuras 6 e 7.

Figura 6 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com instalações avícolas, considerando-se o modelo IR

Figura 7 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com instalações avícolas, considerando-se o modelo F

Tabela 6 - Decomposição dos números de graus de liberdade para o experimento com instalações avícolas

Fontes de variação	gl
Blocos	14
Parcelas[Blocos]	45
Telhas	1
Forros	1
Telhas#Forros	1
Resíduo (A)	42
Horas	3
Blocos#Horas	42
Parcelas[Blocos]#Horas	135
Telhas#Horas	3
Forros#Horas	3
Telhas#Forros#Horas	3
Resíduo (B)	126

Com o auxílio da Tabela 6 concluiu-se que os efeitos de Telhas, Forros e Telhas#Forros estão parcialmente confundidos com o efeito de Parcelas[Blocos]. E os efeitos de Telhas#Horas, Forros#Horas e Telhas#Forros#Horas estão parcialmente confundidos com o efeito de Parcelas[Blocos]#Horas.

Como não existem efeitos totalmente confundidos, para a obtenção dos modelos GR e GF foram incluídas apenas tendências sistemáticas com relação à Horas em GF e covariâncias distintas, também com relação à Horas, em GR. Logo, os modelos GR e F foram descritos por:

$$\text{GR : } \text{Blocos} / \text{uc}(\text{Horas}) + \text{Blocos} \wedge \text{Parcelas} / \text{uc}(\text{Horas}),$$

$$\text{GF : } \text{Telhas} * \text{Forros} * \text{td}(\text{Horas}).$$

A variância de uma observação é dada por:

$$\sigma_B^2 + \sigma_{BP}^2 + \sigma_{BH}^2 + \sigma_{BPH}^2,$$

em que, cada σ^2 denota uma particular fonte de variação, proveniente de fatores cujas letras iniciais estão em subscrito.

As covariâncias entre duas observações realizadas em diferentes horas, mas na mesma parcela é dada por:

$$\sigma_B^2 + \sigma_{BP}^2 + \sigma_{BH}^2 \rho_{hh'} + \sigma_{BPH}^2 \rho'_{hh'} ,$$

em que $\rho_{hh'}$ denota a correlação entre observações do mesmo bloco nas horas h e h' e $\rho'_{hh'}$ denota a correlação entre observações na mesma parcela nas horas h e h' .

A covariância entre observações realizadas em diferentes horas, em diferentes parcelas, porém no mesmo bloco é dada por:

$$\sigma_B^2 + \sigma_{BH}^2 \rho_{hh'} .$$

E a covariância será nula, caso contrário.

Considerando-se os modelos HR e F, construiu-se o modelo Base (ANOVA),

$$\begin{aligned} \text{Telhas*Forros*Horas} \quad | \quad \text{Blocos/ Horas} + \\ + \text{Blocos}\wedge\text{Parcelas}, \end{aligned} \tag{36}$$

para o qual os termos antes de $|$ são de efeito fixo e os termos após são de efeito aleatório.

Para supor algumas tendências com relação aos níveis do fator Horas e também a respeito da variabilidade das observações, foram construídos os gráficos de perfis, apresentados na figura a seguir.

Figura 8 - Gráfico de perfis individuais e médios para o experimento com instalações avícolas

Como se pode notar na Figura 8, existe uma tendência quadrática da temperatura de bulbo seco com relação às horas, desse modo foram consideradas tendências quadráticas, e para captar uma possível falta de ajuste foram introduzidos termos de suavização spline cúbica, como termos de tendências variando aleatoriamente. Sendo assim, o próximo modelo ajustado, ou seja, o modelo reparametrizado, foi:

$$\begin{aligned}
 & \text{Telhas*Forros*lin(Horas)} \mid \\
 & \text{spl(hora)/(Telhas*Forros)} + \text{dev(hora)/(Telhas*Forros)} + \\
 & \text{Blocos}^{\wedge}(\text{lin(Horas)} + \text{spl(hora)} + \text{Horas} + \text{pol(hora, 2)}) + \\
 & \text{Blocos}^{\wedge}\text{Parcelas}^{\wedge}(\text{lin(Horas)} + \text{spl(hora)} + \text{Horas} + \text{pol(hora,2)}),
 \end{aligned} \tag{37}$$

em que, para o termo lin(Horas) , dos termos de tendências variando aleatoriamente, foram considerados componentes aleatórios no intercepto e no coeficiente angular da reta, entre-

tanto, sem correlação entre os mesmos, e hora representa o fator quantitativo para o fator longitudinal Horas.

Os componentes de variância para os termos: $\text{spl}(\text{hora})$, $\text{spl}(\text{hora}) \wedge \text{Telhas}$, $\text{dev}(\text{hora})$, $\text{dev}(\text{hora}) \wedge \text{Telhas}$, $\text{dev}(\text{hora}) \wedge \text{Forros}$, $\text{dev}(\text{hora}) \wedge \text{Telhas} \wedge \text{Forros}$, $\text{Blocos} \wedge \text{Parcelas} \wedge \text{spl}(\text{hora})$ e também para o intercepto dos termos $\text{Blocos} \wedge \text{lin}(\text{Horas})$ e $\text{Blocos} \wedge \text{Parcelas} \wedge \text{lin}(\text{Horas})$, foram estimados na fronteira do espaço paramétrico, portanto foram excluídos do modelo 37, e o modelo reparametrizado foi dado conforme segue:

$$\begin{aligned} & \text{Telhas} * \text{Forros} * \text{lin}(\text{Horas}) | \\ & \text{spl}(\text{hora}) \wedge (\text{Telhas} / \text{Forros}) + \\ & \text{Blocos} \wedge (\text{lin}(\text{Horas}) + \text{spl}(\text{hora}) + \text{Horas} + \text{pol}(\text{hora}, 2)) + \\ & \text{Blocos} \wedge \text{Parcelas} \wedge (\text{lin}(\text{Horas}) + \text{pol}(\text{hora}, 2)), \end{aligned} \tag{38}$$

para o qual os termos $\text{lin}(\text{Horas})$ indicam a presença de parâmetros aleatórios apenas no coeficiente angular das retas em questão.

As estatísticas produzidas para os modelos Base (modelo 36) e reparametrizado (modelo 38), bem como para os demais modelos propostos, são apresentadas na Tabela 7, na qual também estão as estatísticas para os testes de razão de verossimilhanças entre os modelos de interesse.

Inicialmente, tomou-se como modelo de referência o modelo reparametrizado (38) e a inclusão e exclusão de termos no modelo de referência são denotadas na Tabela 7 pelos sinais “+” e “-”, respectivamente. O símbolo ^A indica que a mudança proposta foi aceita e o modelo em questão passou a ser o modelo de referência para as próximas comparações.

Seja um exemplo a primeira modificação proposta no modelo reparametrizado, ou seja, a exclusão do termo $\text{Blocos} \wedge \text{Parcelas} \wedge \text{pol}(\text{hora}, 2)$, uma vez que seu componente de variância associado foi estimado próximo a fronteira do espaço paramétrico ($\sigma_{BP\text{pol}(\text{hora}, 2)}^2 = 0,08$ com erro padrão 0,02). Tal modificação é denotada na tabela por - BP $\text{pol}(\text{hora}, 2)$, a qual não foi aceita, pois seu valor-*p* foi inferior a 0,05, e portanto, o modelo reparametrizado permaneceu como modelo de referência.

Tabela 7 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com instalações avícolas

Modelo	AIC	BIC	REMLRT	gl	valor- <i>p</i>
Base	210,7558	224,4024			
Reparametrizado	173,8651	204,8857			
Modelando $\mathbf{G}$					
- BP pol(hora, 2)	198,8859	226,4598	27,0208	1	< 0,0001
- spl(hora)FT	176,8317	204,4056	4,9667	1	0,0129
- BP lin(Hora)	174,2243	201,7982	2,3593	1	0,0623 ^A
- B pol(hora, 2)	174,2974	198,4245	2,0730	1	0,0750 ^A
- B spl(hora)	173,9467	194,6271	1,6493	1	0,0995 ^A
- spl(hora) F	188,5683	205,802	16,6216	1	< 0,0001
Modelando $\mathbf{\Sigma}$					
+ AR(1)(H)	165,1699	185,8503 ^A			
+ ID _H (H)	161,6075	192,6281	9,5624	3	0,0227 ^A
+ AR _H (1)(H)	163,3517	197,8191	0,2558	1	0,6130
+ us(H)	161,8104	210,0647	9,7971	5	0,0812

^A indica que proposta de modificação do modelo de referência aceita

O modelo resultante para os termos de tendência variando aleatoriamente, foi dado por:

$$\begin{aligned} & \text{Telhas*Forros*lin(Horas)} \mid \text{spl(hora)\wedge Forros/ Telhas} + \\ & + \text{Blocos}\wedge(\text{lin(Horas)} + \text{Horas}) + \\ & + \text{Blocos}\wedge\text{Parcelas}\wedge\text{pol(hora, 2)}, \end{aligned}$$

para o qual o termo lin(hora) corresponde ao componente aleatório no coeficiente angular da reta para Blocos com relação ao níveis do fator Horas. Modelada a estrutura para a matriz de variâncias e covariâncias residuais, $\mathbf{\Sigma}$, o modelo resultante foi:

$$\begin{aligned} & \text{Telhas*Forros*lin(Horas)} \mid \text{spl(hora)\wedge Forros/ Telhas} + \\ & + \text{Blocos}\wedge(\text{lin(Horas)} + \text{Horas}) + \\ & + \text{Blocos}\wedge\text{Parcelas}\wedge(\text{pol(hora, 2)} + \text{idh(Horas)}). \end{aligned} \quad (39)$$

Em seguida, foram verificadas as significâncias dos parâmetros de efeito fixo, e as estatísticas para o teste Wald- F são apresentadas a seguir, salientando que o nível de significância adotado foi 5%. Dois termos foram não significativos, interceptos distintos para a combinação entre níveis de Telhas e Forros (Telhas#Forros) e, coeficientes angulares distintos para níveis de Forros (Forros#hora).

Tabela 8 - Testes de Wald- F para os efeitos fixos presentes no modelo 39, do experimento com instalações avícolas

Fontes de variação	gl	gl den	Wald- F	valor- p
Intercepto	1	55,7	2691,0000	< 0,0001
Telhas	1	89,2	20,1500	< 0,0001
Forros	1	96,5	4,1350	0,0463
hora	1	22,1	89,6400	< 0,0001
Telhas#Forros	1	89,2	0,8693	0,3546
Telhas#hora	1	115,8	7,3450	0,0082
Forros#hora	1	91,3	2,4470	0,1239
Telhas#Forros#hora	1	115,8	6,0210	0,0164

O modelo contendo os termos fixos significativos foi ajustado, um novo teste Wald- F foi realizado e o termo Telhas#Forros#hora foi considerado não significativo. Portanto, o modelo resultante foi dado por:

$$\begin{aligned}
 \text{Forros} + \text{Telhas} * \text{lin}(\text{Horas}) \quad | \quad \text{spl}(\text{hora}) \wedge \text{Forros} / \text{Telhas} + \quad (40) \\
 + \text{Blocos} \wedge (\text{lin}(\text{Horas}) + \text{Horas}) + \\
 + \text{Blocos} \wedge \text{Parcelas} \wedge (\text{pol}(\text{hora}, 2) + \text{idh}(\text{Horas})).
 \end{aligned}$$

Conforme a proposta do método de seleção de modelos adotado, a importância de outros termos foi verificada, porém nenhuma mudança foi aceita. Os critérios de informação para o referido modelo são 152,931 e 184,067, para AIC e BIC, respectivamente.

Os gráficos de dispersão para o EBLUP e para os resíduos condicionais estudantizados são apresentados nas Figuras 9 e 10, respectivamente. Na primeira pode-se verificar uma maior dispersão para os valores iniciais e finais do EBLUP, sendo este um indicativo de mau ajuste do modelo. Na segunda figura tem-se a presença de dois pontos atípicos.

Figura 9 - Gráfico dos valores para o EBLUP (modelo 40) para o experimento com instalações avícolas

Figura 10 - Gráfico dos resíduos condicionais estudatizados (modelo 40) para o experimento com instalações avícolas

Nas Figuras 11 e 12 são apresentados, respectivamente, o EBLUP com confundimento mínimo e os resíduos com confundimento mínimo. Pode-se notar que ambos não satisfazem a pressuposição de normalidade.

Figura 11 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 40) para o experimento com instalações avícolas

Figura 12 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 40) para o experimento com instalações avícolas

Uma vez que a pressuposição de normalidade não foi satisfeita, novos modelos

foram propostos, e o que melhor se adequou ao conjunto de dados foi o modelo

$$\begin{aligned} \text{Forros} + \text{Telhas} * \text{lin}(\text{Horas}) \quad | \quad \text{spl}(\text{hora}) + \\ + \text{Blocos} \wedge (\text{lin}(\text{Horas}) + \text{pol}(\text{hora}, 2) + \text{Horas}) + \\ + \text{Blocos} \wedge \text{Parcelas} \wedge \text{idh}(\text{Horas}). \end{aligned} \quad (41)$$

Embora o modelo 41 tenha apresentado maiores valores para os critérios AIC e BIC, 172,6364 e 200,3131, este apresentou EBLUP e resíduos mais próximos da normalidade, porém ainda com vários pontos fora do intervalo limitado pelo envelope, como pode ser verificado nas Figuras 15 e 16. O modelo apresentou também EBLUP disperso aleatoriamente ao redor de zero (Figura 13). Em contrapartida, no gráfico de dispersão para os resíduos condicionais estudatizados (Figura 14), há quatro pontos atípicos.

Figura 13 - Gráfico dos valores para o EBLUP (modelo 41) para o experimento com instalações avícolas

Figura 14 - Gráfico dos resíduos condicionais estudatizados (modelo 41) para o experimento com instalações avícolas

Figura 15 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 41) para o experimento com instalações avícolas

Figura 16 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 41) para o experimento com instalações avícolas

A seguir são apresentadas as curvas ajustadas para o modelo selecionado (41), a partir das quais pode-se concluir que coberturas com telhas de cerâmica e forro A retém mais calor, pois apresentaram maior temperatura de bulbo seco. E, coberturas com telhas de fibrocimento e forro B retém menos calor, sendo esta última a mais indicada para cobertura de instalações avícolas.

Figura 17 - Curvas ajustadas de acordo com o modelo 41 para o experimento com instalações avícolas

4.2 Experimento com polpas de Eucalipto

Inicialmente, definiu-se cada uma das camadas do experimento com polpas de Eucalipto, listadas a seguir:

- (i) A primeira fase do experimento deu origem à primeira camada, a qual contém os fatores {Espécies, Idades, Lotes, Grupos}, com a seguinte fórmula intra-camada:

$$(2 \text{ Espécies} * 2 \text{ Idades}) / 3 \text{ Lotes} / 4 \text{ Grupos.} \quad (42)$$

- (ii) Da segunda fase do experimento, originou-se a segunda camada, contendo os fatores {Cozimentos, Amostras}, tal que:

$$48 \text{ Cozimentos} / 6 \text{ Amostras.} \quad (43)$$

- (iii) A terceira fase deu origem a duas camadas, uma com os fatores {Rodadas, Posições}, tal que:

$$48 \text{ Rodadas} / 6 \text{ Posições}, \quad (44)$$

e outra com o fator {Tempos}, com

$$6 \text{ Tempos}. \quad (45)$$

Como notação, foram utilizadas as letras iniciais dos fatores, E(spécies), I(dades), L(otes), G(rupos) e T(empos).

Deveriam existir aleatorizações dos Grupos aos Cozimentos e dos Cozimentos às Rodadas, como ilustrado no diagrama de aleatorização, Figura 18. Rodadas e Posições (dentro de Rodadas) são unidades experimentais, pois possuem fatores aleatorizados a eles, e a aleatorização foi restrita às posições dentro de rodadas como indicado pela relação de aninhamento.

Figura 18 - Diagrama de aleatorização para o experimento com polpas de Eucalipto

Nesse experimento há apenas um fator longitudinal, Tempos. A fórmula intracamada 44 corresponde às unidades observacionais, Posições[Rodadas], e por isso foram incluídas no modelo IR, bem como a fórmula 43. As fórmulas 42 e 45 não recebem aleatorizações, apenas são aleatorizados a Cozimentos e Posições, respectivamente, e sendo assim foram incluídas no modelo IF. Logo,

$$\text{IR : } \text{Cozimentos} / \text{Amostras} + \text{Rodadas} / \text{Posições},$$

$$\text{IF : } (\text{Espécies} * \text{Idades}) / \text{Lotes} / \text{Grupos} + \text{Tempos}.$$

Um segundo passo foi considerar interações de interesse prático como Espécies#Tempos, Idades#Tempos e Espécies#Idades#Tempos. Assim, ao invés de se considerar a relação aditiva entre as fórmulas para as camadas de grupos e de tempos, considerou-se

a relação cruzada. Foram assumidos como de efeito fixo os fatores Espécies, Idades e Tempos, os demais foram considerados aleatórios, compondo os seguintes modelos HR e F:

$$\text{HR : } \text{Cozimentos} / \text{Amostras} + \text{Rodadas} / \text{Posições} + (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes}) / \text{Tempos} + \\ + (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}) / \text{Tempos},$$

$$\text{F : } \text{Espécies} * \text{Idades} * \text{Tempos}.$$

Com o objetivo de compreender melhor os confundimentos presentes nesse experimento, foi construída a tabela de decomposição dos números de graus de liberdade, apresentada na Tabela 9. Para auxiliar na obtenção dos graus de liberdade referentes a cada efeito, foram construídos os diagramas de Hasse para cada fase do experimento, apresentados nas Figuras 19 e 20.

Figura 19 - Diagramas de Hasse para obtenção dos números de graus de liberdade para o experimento com polpas de Eucalipto, considerando-se a primeira fase (camada das amostras) e a segunda fase (camada das posições) do experimento

Figura 20 - Diagrama de Hasse para obtenção dos números de graus de liberdade para o experimento com polpas de Eucalipto, considerando-se a terceira fase do experimento (camadas dos grupos e dos tempos)

Tabela 9 - Decomposição dos números de graus de liberdade para o experimento com polpas de Eucalipto

Camada: Posições		Camada: Amostras		Camadas: Grupos-Tempos	
Fontes de variação	gl	Fontes de variação	gl	Fontes de variação	gl
Rodadas	47	Cozimentos	47	Espécies	1
				Idades	1
				Espécies#Idades	1
				Lotes[E∧I]	8
				Grupos[E∧I∧L]	36
Posições[Rodadas]	240	Amostras[Cozimentos]	240	Tempos	5
				Espécies#Tempos	5
				Idades#Tempos	5
				Espécies#Idades#Tempos	5
				Lotes[E∧I]#Tempos	40
				Grupos[E∧I∧L]#Tempos	180

Com o auxílio da Tabela 9, concluiu-se que o efeito de Cozimentos está totalmente confundido com o efeito de Rodadas e com os efeitos provenientes da fórmula (Espécies*Idade)/Lotes/Grupos e, o efeito de Amostras[Cozimentos] está totalmente confundido com o efeito de Posições[Rodadas]. Sendo assim, os efeitos de Cozimentos, Rodadas e Amostras[Cozimentos] podem ser excluídos do modelo. Portanto, GR e GF foram descritos por:

$$\begin{aligned} \text{GR :} \quad & \text{Rodadas} \wedge \text{Posições} + (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes}) / \text{uc}(\text{Tempos}) + \\ & + (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}) / \text{uc}(\text{Tempos}) \\ \text{GF :} \quad & \text{Espécies} * \text{Idades} * \text{td}(\text{Tempos}), \end{aligned}$$

nos quais o termo Rodadas $\wedge$ Posições corresponde a matriz Σ .

Vale ressaltar que a variância de uma observação é dada por:

$$(\sigma_{EILG}^2 + \sigma_C^2 + \sigma_R^2) + (\sigma_{CA}^2 + \sigma_{RP}^2) + (\sigma_{EILGT}^2 + \sigma_{CT}^2 + \sigma_{RT}^2) + \sigma_{EIL}^2 + \sigma_{EILT}^2,$$

em que, cada σ^2 denota uma particular fonte de variabilidade proveniente de uma particular combinação de fatores cujas letras iniciais são dadas em subscrito. Componentes de variância entre parênteses não podem ser estimados separadamente.

A covariância entre duas medidas realizadas em diferentes tempos, porém, na mesma rodada é dada por:

$$(\sigma_{EILG}^2 + \sigma_C^2 + \sigma_R^2) + (\sigma_{CA}^2 + \sigma_{RP}^2) + (\sigma_{EILGT}^2 + \sigma_{CT}^2 + \sigma_{RT}^2)\rho_{tt'} + \sigma_{EIL}^2 + \sigma_{EILT}^2\rho'_{tt'},$$

em que, $\rho_{tt'}$ denota a correlação entre observações da mesma rodada nos tempos t e t' , e $\rho'_{tt'}$ denota a correlação entre observações do mesmo lote nos tempos t e t' .

E, a covariância entre observações de diferentes níveis de Rodadas foi dada por:

$$\sigma_{EIL}^2 + \sigma_{EILT}^2\rho'_{tt'},$$

para observações no mesmo lote e grupos diferentes, e zero, caso contrário.

A partir do modelo composto por F e HR, tal que, de HR foram excluídos os termos totalmente confundidos com algum já presente no modelo, foi construído o modelo Base (ANOVA),

$$\begin{aligned} \text{Espécies} * \text{Idades} * \text{Tempo} \quad | \quad & (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes}) / \text{Tempos} + \\ & + (\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}) \end{aligned} \quad (46)$$

em que, os termos antes de | são de efeito fixo e os termos após são de efeito aleatório.

Para dar continuidade às análises foram observados os gráficos de perfis para cada indivíduo, apresentados na Figura 21, na qual pode-se notar uma suave curvatura ao longo do tempo e uma possível heterogeneidade de variâncias, principalmente, entre Tempos.

Figura 21 - Gráfico de perfis individuais e médios para o experimento com polpas de Eucalipto

Para examinar o comportamento da variável resposta ao longo do tempo foram considerados modelos de suavização spline cúbica. E, como na Figura 21 existem evidências de variações aleatórias, modelos de tendências variando aleatoriamente foram ajustados. O

próximo modelo ajustado foi:

$$\begin{aligned} & \text{Espécies*Idades*lin(Tempos)} | \\ & \text{spl(tempo)/(Espécies*Idades) + dev(tempo)/(Espécies*Idades) +} \\ & \text{Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge(\text{lin(Tempos) + spl(tempo) + Tempos) +} \\ & \text{Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge\text{Grupos}\wedge(\text{lin(Tempos) + spl(tempo) + Tempos}), \end{aligned} \quad (47)$$

em que para o termo lin(Tempos) , dos termos de tendências variando aleatoriamente, foram considerados componentes aleatórios no intercepto e no coeficiente angular da reta, porém, sem correlação entre os mesmos, e tempo representa o fator quantitativo para o fator longitudinal Tempos.

Todos os termos estimados na fronteira do espaço paramétrico, ou seja, $\text{spl(tempo)}\wedge\text{Espécies}$, $\text{spl(tempo)}\wedge\text{Espécies}\wedge\text{Idades}$, dev(tempo) , $\text{dev(tempo)}\wedge\text{Espécies}$, $\text{dev(tempo)}\wedge\text{Espécies}\wedge\text{Idades}$, $\text{Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge\text{Tempos}$ e $\text{Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge\text{spl(tempo)}$, foram excluídos do modelo 47, ficando o modelo reparametrizado da seguinte forma:

$$\begin{aligned} & \text{Espécies*Idades*lin(Tempos)} | \\ & \text{spl(tempo)/Idades + dev(tempo)}\wedge\text{Idades + Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge\text{lin(Tempos) +} \\ & \text{Espécies}\wedge\text{Idades}\wedge\text{Lotes}\wedge\text{Grupos}\wedge(\text{lin(Tempos) + spl(tempo) + Tempos}), \end{aligned} \quad (48)$$

produzindo as estatísticas descritas na Tabela 10 para o modelo reparametrizado.

Os resultados para as comparações dos modelos de interesse também podem ser observados na Tabela 10. Os dois primeiros modelos são, Base (ANOVA), apresentado no modelo 46 e o modelo reparametrizado, apresentado no modelo 48.

Na referida tabela os sinais “+” e “-” em frente as iniciais dos fatores, ou função dos mesmos, indicam que os termos foram incluídos e excluídos, respectivamente, do modelo em questão, sempre com base no modelo de referência, que inicialmente tratou-se do modelo reparametrizado, porém, com relação a cada um dos modelos seguintes, deu-se pelo último modelo que teve a mudança proposta (inclusão ou exclusão de termos, modificações nas estruturas de variâncias e covariâncias, etc.) aceita, ou seja, o modelo que possui um ^A em seu valor-*p*, considerando-se um nível de 5% de significância.

Como exemplo, considerando o termo $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} \wedge \text{spl}(\text{tempo})$, o componente de variância associado a esse termo foi estimado próximo à fronteira do espaço paramétrico, $\hat{\sigma}_{EILG_{spl}(T)}^2 = 4,11 \times 10^{-7}$, com erro padrão igual à $2,22 \times 10^{-6}$. Tal termo foi excluído do próximo modelo ajustado, denotado por - EILG spl(t).

O modelo - EILG spl(t) foi testado, para o qual o modelo de referência foi o modelo 48. Como o valor- p foi igual a 0,42675, o A indica que a mudança foi aceita, ou seja, o modelo - EILG spl(t) será o próximo modelo de referência.

Seguindo o mesmo critério, o próximo modelo exclui o termo $\text{spl}(\text{tempo}) \wedge \text{Idades}$, com componente de variância estimado dado por $\sigma_{spl(t)I}^2 = 1,79 \times 10^{-6}$ com erro padrão $3,47 \times 10^{-6}$. Como o valor- p para o teste da razão de verossimilhanças foi superior a 0,05, decidiu-se por remover o termo do modelo, e o modelo de referência passou a ser - spl(tempo) \wedge Idades.

Procedeu-se de modo análogo a análise dos demais termos apresentados na Tabela 10, salientando que o modelo - EILG int-coef + EILG componente exclui o termo de tendência linear variando aleatoriamente para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}$ (com covariância nula entre intercepto e coeficiente angular) e inclui um componente aleatório para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}$. O modelo EILG cov int-coef inclui um parâmetro de covariância entre o intercepto e o coeficiente angular para o termo de tendência linear variando aleatoriamente para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}$.

Para as diferentes estruturas para a matriz de variâncias e covariâncias Σ , o modelo - EILG cov int-coef + $ID_H(T)$ denota a exclusão do termo de covariância entre o intercepto e o coeficiente angular do modelo de tendência linear variando aleatoriamente para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}$ e inclui variâncias diferentes para tempos diferentes. Tal exclusão foi considerada, pois a permanência do termo no modelo criou uma dependência linear no modelo com estrutura ID_H para Tempos. E, os modelos + T cov diferentes para L e + T cov diferentes para EI, permitem estimar variâncias residuais distintas para cada nível de Lotes e de $\text{Espécies} \wedge \text{Idades}$, respectivamente.

A Tabela 10 apresentada a seguir contém as estatísticas AIC, BIC, a razão das verossimilhanças, a diferença entre os números de graus de liberdade dos modelos de referência e aninhado e o valor- p , para cada modelo estudado.

Tabela 10 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com polpa de Eucalipto

Modelo	AIC	BIC	REMLRT	gl	valor- <i>p</i>
Base	-1840,398	-1826,094			
Reparametrizado	-2008,815	-1976,102			
Modelando G					
- EILG $\wedge$ spl(tempo)	-2010,781	-1981,703	0,0341	1	0,4268 ^A
- spl(tempo) $\wedge$ Idades	-2012,028	-1986,585	0,7527	1	0,1928 ^A
- spl(tempo)	-2010,238	-1988,429	3,7902	1	0,0258
- dev(tempo) $\wedge$ Idades	-2010,494	-1988,685	3,5339	1	0,0301
- EILG int-coef + EILG componente	-1969,329	-1947,520	44,699	1	< 0,0001
+ EILG cov int-coef	-2015,497	-1986,418	5,4684	1	0,0194 ^A
- dev(tempo) $\wedge$ Idades	-2013,318	-1987,874	4,1788	1	0,0205
Modelando Σ					
+ AR(1)(T)	-2014,032	-1981,319	0,5360	1	0,4641
- EILG cov int-coef + ID _H (T)	-2060,483	-2016,822	52,9862	4	< 0,0001 ^A
+ AR _H (1)(T)	-2043,625	-1996,373	-14,8578	1	1
+ T cov diferentes para L	-2038,479	-1951,244	1,9967	12	0,9994
+ T cov diferentes para EI	-2044,511	-1935,467	20,0280	18	0,3312
- EILG int-coef + EILG componente	-2025,201	-1985,219	37,2813	1	< 0,0001
- spl(tempo)	-2059,194	-2019,172	3,2888	1	0,0698

^A indica que proposta de modificação do modelo de referência aceita

Verificadas as significâncias dos termos de tendência variando aleatoriamente, o modelo resultante foi:

$$\begin{aligned}
 & \text{Espécies} * \text{Idades} * \text{lin}(\text{Tempos}) \quad | \quad \text{spl}(\text{tempo}) + \text{dev}(\text{tempo}) \wedge \text{Idades} + \\
 & + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{lin}(\text{Tempos}) + \\
 & + (\text{Cozimentos} + \text{Rodadas} + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}) \wedge \text{lin}(\text{Tempos}),
 \end{aligned}$$

para o qual foram considerados coeficientes aleatórios no intercepto e no coeficiente angular para os termos de tendência linear variando aleatoriamente, com covariância nula para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes}$ e covariância não nula para $\text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos}$. Em seguida, modeladas as covariâncias entre os tempos, ou seja, a matriz Σ , o modelo resultante

foi dado por:

$$\begin{aligned} \text{Espécies*Idades*lin(Tempos)} \quad | \quad \text{dev(tempo)} \wedge \text{Idades} + \\ + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{lin(Tempos)} + \\ + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} \wedge (\text{lin(Tempos)} + \text{idh(Tempos)}). \end{aligned} \quad (49)$$

Os resultados para o teste Wald- F para o modelo 49 são apresentados na Tabela 11, em que foram significativos a 5%, apenas o intercepto comum e o coeficiente angular comum.

Tabela 11 - Testes de Wald- F para os efeitos fixos presentes no modelo 49, do experimento com polpas de Eucalipto

Fontes de variação	gl	gl den	Wald- F	valor- p
Intercepto	1	13,6	$1,265 \times 10^4$	$< 0,0001$
Espécies	1	8,0	4,798	0,0599
Idades	1	12,5	2,507	0,1373
tempo	1	14,4	$3,192 \times 10^2$	$< 0,0001$
Espécies#Idades	1	8,0	$4,673 \times 10^{-2}$	0,8343
Espécies#tempo	1	268,0	2,354	0,1288
Idades#tempo	1	268,0	2,816	0,0968
Especies#Idades#tempo	1	268,0	$2,447 \times 10^{-1}$	0,6166

Entretanto, decidiu-se, baseado em 10% de significância, que os termos Espécies e Idades $\wedge$ tempo deveriam permanecer no modelo, dado por:

$$\begin{aligned} \text{Espécies} + \text{lin(Tempos)}/\text{Idades} \quad | \quad \text{dev(tempo)} \wedge \text{Idades} + \\ + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{lin(Tempos)} + \\ + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} \wedge (\text{lin(Tempos)} + \text{idh(Tempos)}). \end{aligned} \quad (50)$$

Um novo teste Wald- F foi realizado, no qual todos os termos fixos do modelo 50 foram significativos a 5%, conforme apresentado na tabela a seguir.

Tabela 12 - Testes de Wald F para os efeitos fixos presentes no modelo 50, do experimento com polpas de Eucalipto

Fontes de variação	gl	gl den	Wald- F	valor- p
Intercepto	1	15,9	$1,387 \times 10^4$	$< 0,0001$
Espécies	1	9,4	5,459	0,0443
tempo	1	16,6	$3,105 \times 10^2$	$< 0,0001$
tempo#Idades	1	272,0	4,736	0,0316

Novas estruturas para a matriz $\mathbf{G}$ foram propostas, conforme pode ser verificado na Tabela 13, e o modelo selecionado foi dado conforme segue:

$$\begin{aligned}
 &\text{Espécies} + \text{lin(Tempos)/Idades} \quad | \quad \text{spl(tempo)} \wedge \text{Idades} + \\
 &+ \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{lin(Tempos)} + \\
 &+ \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} \wedge (\text{lin(Tempos)} + \text{idh(Tempos)}).
 \end{aligned} \tag{51}$$

Tabela 13 - Critérios de informação e testes da razão de verossimilhanças para os modelos ajustados para o experimento com polpa de Eucalipto

Modelo	AIC	BIC	REMLRT	gl	valor- p
Modelo 50	-2092,084	-2051,945			
+ spl(tempo)∧Idades	-2093,041	-2049,253	2,957134	1	0,0427 ^A
- dev(tempo)∧Idades	-2094,791	-2054,652	0,2500605	1	0,3085 ^A
- EILG int-coef + EILG componente	-2076,918	-2040,428	19,87302	1	$4,1380 \times 10^{-6}$

^A indica que proposta de modificação do modelo de referência aceita

Nas Figuras 22 e 23 são apresentados os gráficos de dispersão dos valores obtidos para o EBLUP ($\hat{\mathbf{b}}$) e para o resíduo condicional estudentizado ($\mathbf{r}_c^{\text{studentizado}}$), respectivamente, considerando-se o modelo ajustado 51. Na Figura 22 pode-se notar que os pontos estão aleatoriamente distribuídos ao redor de zero, o mesmo acontece quando observada a Figura 23, entretanto nesta última, há a presença de dois pontos atípicos.

Figura 22 - Gráfico dos valores para o EBLUP (modelo 51) para o experimento com polpas de Eucalipto

Figura 23 - Gráfico dos resíduos condicionais estudatizados (modelo 51) para o experimento com polpas de Eucalipto

Conforme pode ser comprovado pelas Figuras 24 e 25, existem muito pontos fora do intervalo de simulação e, portanto, embora o modelo 51 tenha apresentado melhores critérios de informação e tenha sido selecionado pelo teste da razão de verossimilhanças, o mesmo não é adequado, pois não atende às pressuposições de normalidade dos modelos

lineares mistos.

Figura 24 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 51) para o experimento com polpas de Eucalipto

Figura 25 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 51) para o experimento com polpas de Eucalipto

Em busca de um modelo parcimonioso que atendesse as pressuposições de normalidade, tanto dos resíduos quanto do vetor de parâmetros de efeito aleatório, vários modelos foram estudados, e o modelo que apresentou melhores resultados quanto à distribuição

do EBLUP com confundimento mínimo e dos resíduos com confundimento mínimo, foi o modelo:

$$\begin{aligned}
 & (\text{Espécies} + \text{Idades}) * \text{lin}(\text{Tempos}) \quad | \quad \text{spl}(\text{tempo}) \wedge \text{Idades} + \\
 & + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} + \\
 & + \text{Espécies} \wedge \text{Idades} \wedge \text{Lotes} \wedge \text{Grupos} \wedge \text{Idh}(\text{Tempos}).
 \end{aligned} \tag{52}$$

As estatísticas e resultados para o teste Wald- F , para o modelo 52, são apresentados na tabela a seguir, na qual tem-se todos os termos altamente significativos.

Tabela 14 - Testes de Wald- F para os efeitos fixos presentes no modelo 52, do experimento com polpas de Eucalipto

Fontes de variação	gl	gl den	Wald- F	valor- p
Intercepto	1	76,2	67270,00	< 0,0001
Espécies	1	44,8	31,77	< 0,0001
Idades	1	44,9	18,30	< 0,0001
tempo	1	94,7	1909,00	< 0,0001
Espécies#tempo	1	273,0	11,13	0,001
Idades#tempo	1	273,0	16,99	< 0,0001

E, os gráficos para diagnósticos, em que, o EBLUP e os resíduos condicionais estudatizados estão aleatoriamente distribuídos ao redor de zero, Figuras 26 e 27, respectivamente. Embora vários pontos estejam fora do intervalo para o envelope de simulação, considerando-se os resíduos com confundimento mínimo (Figura 29), este modelo proporcionou um melhor ajuste quando observado o gráfico de quantil-quantil com envelope de simulação para o EBLUP com confundimento mínimo (Figura 28).

Figura 26 - Gráfico dos valores para o EBLUP (modelo 52) para o experimento com polpas de Eucalipto

Figura 27 - Gráfico dos resíduos condicionais estudatizados (modelo 52) para o experimento com polpas de Eucalipto

Figura 28 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 52) para o experimento com polpas de Eucalipto

Figura 29 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 52) para o experimento com polpas de Eucalipto

Apesar do modelo 52 ter apresentado maiores valores para os critérios de informação AIC e BIC, -2005,927 e -1976,792, respectivamente, quando comparados com o modelo 51, o mesmo apresentou melhores ajustes quanto a distribuição dos vetores de efeito aleatório, e portanto, optou-se pelo modelo 52 como modelo final.

As curvas ajustadas para o modelo selecionado são apresentadas a seguir, com as quais pode-se concluir que polpas de árvores da espécie *E. alba* com cinco anos de idade produziram folhas de papel com maior peso específico aparente, enquanto que polpas das árvores da espécie *E. saligna* com sete anos, produziram folhas de papel com menor peso específico aparente.

Figura 30 - Curvas ajustadas de acordo com o modelo 52 para o experimento com polpas de Eucalipto

Devido a proximidade entre as curvas ajustadas para as espécies *E. alba* com 5 anos e *E. saligna* com 7 anos, novos ajustes foram realizados, entretanto considerando-se um novo fator, denominado Tratamentos, em que, um primeiro nível foi atribuído àquelas unidades que pertenciam a espécie *E. alba* e tinham 5 anos ou pertenciam a espécie *E. saligna* e tinham 7 anos, o segundo nível foi atribuído às unidades que pertenciam a espécie *E. alba* e tinham 7 anos e o terceiro nível, atribuído às unidades que pertenciam a espécie *E. saligna* e tinham 5 anos.

Procedeu-se a seleção do modelo de modo análogo a anterior, e o modelo final ajustado foi dado por:

$$\begin{aligned} \text{Tratamentos} * \text{lin}(\text{Tempos}) \quad | \quad \text{spl}(\text{tempo}) + \text{Tratamentos} \wedge \text{Lotes} \quad (53) \\ + \text{Tratamentos} \wedge \text{Lotes} \wedge \text{Grupos} \wedge \text{us}(\text{Tempos}), \end{aligned}$$

para o qual todos os parâmetros de efeito fixo foram significativos ao nível de 5% de significância.

Os gráficos de dispersão podem ser observados nas Figuras 31 e 32, nos quais pode-se notar que os EBLUPs e os resíduos condicionais estudatizados estão aleatoriamente distribuídos ao redor de zero. Com relação à distribuição do EBLUP pode-se concluir pela normalidade (Figura 33), porém quando observada a Figura 34, notam-se vários pontos fora do intervalo limitado pelo envelope, entretanto, o modelo 53, foi o modelo que apresentou distribuições mais próximas da normalidade, tanto para os resíduos quanto para o EBLUP.

Figura 31 - Gráfico dos valores para o EBLUP (modelo 53) para o experimento com polpas de Eucalipto

Figura 32 - Gráfico dos resíduos condicionais estudatizados (modelo 53) para o experimento com polpas de Eucalipto

Figura 33 - Gráfico de quantil-quantil para os valores preditos com confundimento mínimo (modelo 53) para o experimento com polpas de Eucalipto

Figura 34 - Gráfico de quantil-quantil para os resíduos com confundimento mínimo (modelo 53) para o experimento com polpas de Eucalipto

As curvas ajustadas se apresentadas na figura a seguir:

Figura 35 - Curvas ajustadas de acordo com o modelo 53 para o experimento com polpas de Eucalipto

Sendo os modelos 52 e 53, modelos não encaixados, a comparação entre os mesmos foi feita utilizando-se os critérios de informação, cujos resultados são apresentados na tabela a seguir:

Tabela 15 - Critérios de informação para comparação dos modelos 52 e 53

	AIC	BIC
Modelo 52	-2005,972	-1976,792
Modelo 53	-2064,627	-1980,863

Como o modelo 53 apresentou menores valores tanto para AIC quanto para BIC, concluiu-se que não existem diferenças significativas entre o peso específico aparente do papel produzido a partir de plantas da espécie *E. alba* com 5 anos e do papel produzido a partir de plantas da espécie *E. saligna* com 7 anos.

Embora as três curvas apresentem comportamentos semelhantes, encontram-se distantes umas das outras. Desse modo, conclui-se que o peso específico aparente do papel produzido a partir de plantas da espécie *E. alba* com 7 anos difere do peso específico do papel produzido a partir de plantas da espécie *E. saligna* com 5 anos. E, o peso específico aparente do papel produzido a partir de plantas das espécies *E. alba* com 5 anos e *E. saligna* com 7 anos difere do peso específico aparente do papel produzido a partir de plantas das espécies *E. alba* com 7 anos e *E. saligna* com 5 anos.

5 CONSIDERAÇÕES FINAIS

A discriminação dos fatores como aleatorizados e não aleatorizados, assim como a definição de cada camada presente no experimento, permitiram a construção dos diagramas de Hasse e de aleatorização, os quais possibilitaram um maior entendimento a respeito do confundimento entre fatores em camadas distintas.

O emprego das técnicas de formulação de modelos mistos facilitou a obtenção de modelos baseados na aleatorização, permitindo que todos os fatores (não inteiramente confundidos) fossem incluídos em um modelo inicial para análise.

O pacote estatístico ASReml-R mostrou-se eficiente para o ajuste de modelos lineares mistos, por possuir um amplo leque de funções para diferentes estruturas considerando-se as matrizes de variâncias e covariâncias.

Os resíduos com confundimento mínimo, assim como o EBLUP com confundimento mínimo, foram de grande auxílio para a seleção de um modelo adequado, uma vez que as pressuposições de normalidade dos resíduos e do vetor de parâmetros de efeito aleatório devem ser atendidas, pressuposições estas verificadas com maior facilidade fazendo uso das funções implementadas neste trabalho.

REFERÊNCIAS

- AKAIKE, H. A New Look at the Statistical Model Identification. **IEEE Transactions on automatic control**, New York, v. 19, n. 6, p. 716-723, Dec. 1974.
- ALCARDE, R. **Fundamentos do diagrama de Hasse e aplicações à experimentação**. 2007. 99 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2008.
- ALLEN, D. M. The Relationship between Variable Selection and Data Augmentation and a Method for Prediction. **Technometrics**, Alexandria, v. 16, n. 1, p. 125-127, Feb 1974.
- BAILEY, R.A. Hasse diagrams in designed experiments: a pictorial aid to thinking about blocking stratification, degrees of freedom, randomization, and analysis of variance. In SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., Londrina, 2005. **Minicurso...** Londrina: UEL, 2005. 96 p.
- BALZARINI, M. Applications of mixed models in plant breeding. In: KANG, M.S. (Ed.) **Quantitative genetics, genomics and plant breeding**. New York: CABI Publishing. 2001. p. 353-363.
- BASFORD, K.E.; FEDERER, W.T.; DELACY, I.H. Mixed Model formulations for Multi-Environment Trials. **Agronomy Journal**, Madison, v. 96, n. 1, p. 143-147, Jan. 2004.
- BATES, D.; MAECHLER, M.; BOLKER, B. **Linear mixed-effects models using S4 classes**, 2011. 34 p.
- BECKMAN, R.J.; NACHTSHEIM, C.J.; COOK, R.D. Diagnostics for mixed-model analysis of variance. **American Society for Quality**, Madison, v. 29, n. 4, p. 413-426, Nov. 1987.
- BRIEN, C.J. Analysis of variance tables based on experimental structure. **Biometrics**, Arlington, v. 39, n. 1, p. 53-59, Mar. 1983.
- BRIEN, C.J. Determining the analysis of variance table. In: - **Statistical Modelling**. Disponível em: <<http://chris.brien.name/ee2/course/SM06.pdf>>. Acesso em: 20 mar. 2007.
- BRIEN, C.J. **Multitiered experiments**. Disponível em: <<http://chris.brien.name/multitier/>>. Acesso em: 17 jul. 2009.
- BRIEN, C.J.; BAILEY, R.A. Multiple randomizations. **Journal of the Royal Statistical Society**. Series B, London, v. 68, n. 4, p. 571-609. 2006.
- BRIEN, C.J.; DEMÉTRIO, C.G.B. Formulating Mixed Models for Experiments, Including Longitudinal Experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 14, n. 3, p. 253-280. 2009.
- BUTLER, D.G.; CULLIS, B.R.; GILMOUR, A.R.; GOGEL, B.J. **Mixed models for S language environments: ASReml-R reference manual**. v.3 Mar. 2009, 149 p.
- CALIŃSKI, S.; CZAJKA, S.; KACZMAREK, Z. KRAJEWSKI, P.; PILARCZYK, W. Analyzing Multi-environment Variety Trials Using Randomization-Derived Mixed Models. **Biometrics**, Arlington, v. 61, n. 2, p. 448-455, June 2005.
- CASANOVA, F.; MACCHIAVELLI, R.; BALZARINI, M. Error Variation in Multi-environment Peanut Trials: Within-Trial Spatial Correlation and Between-Trial Heterogeneity. **Crop Science**, Madison, v. 45, n. 5, p. 1928-1933, Sep. 2005.

CASTRO, A.C.. **Avaliação da eficiência térmica de materiais utilizados como sistemas de cobertura em instalações avícolas**. 2011. 100 p. Dissertação (Mestrado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

CHRISTENSEN, R.; PEARSON, L.M.; JOHNSON, W. Case-Deletion Diagnostics for Mixed Models. **Technometrics**, Alexandria, v. 34, n. 1, p. 38-45, Feb. 1992.

CLAESKENS, G.; HART, J.D. Goodness-of-fit tests in mixed models. **Sociedad de Estadística e Investigación Operativa**, Leuven, v. 18, n. 2, p. 213-239, Feb. 2009.

COOK, R.D. Detection of Influential Observations in Linear Regression. **Technometrics**, Alexandria, v. 19, n. 1, p. 15-18, Feb 1977.

COOK, R.D.; WEISBERG, S. **Residuals and Influence in Regression**. 1. ed. New York: Chapman & Hall, 1982, 230 p.

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 of Statistical Computation and Simulation**, London, v. 54, v. 4, p. 363-378, 1996.

FITZMAURICE, G.; DAVIDIAN, M.; VERBEKE, G.; MOLENBERGHS, G.. **Longitudinal Data Analysis**. 1. ed. Boca Raton: Chapman & Hall, 2009, 618p.

GILMOUR, A.R.; THOMPSON, R.; CULLIS, B.R. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, n. 4, p. 1440-1450, Dec. 1995.

HARVILLE, D.A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. **Journal of the American Statistical Association**, Alexandria, v. 72, n. 358, p. 320-338, Jun 1977.

HARVILLE, D.A. **Matrix Algebra from a Statistician's Perspective**. 1. ed. Secaucus: Springer, 1997, 648p.

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

HENDERSON, C.R. Selection index and expected genetic advance. In: HANSON, W.D.; ROBINSON, H.F. (Ed.). **Statistical genetics and plant breeding**. Washington DC: National Academy of Genetic Advance - National Research Council, 1963. p. 141-163.

HILDEN-MINTON, J.A. **Multilevel Diagnostics for Mixed and Hierarchical Linear Models**. 1995. 110 p. PhD Thesis (Philosophy in Mathematics) - University of California, Los Angeles, 1995.

HÜHN, M. Nonparametric analysis of genotype×environment interactions by ranks. In KANG, M.S.; GAUCH Jr, H.G. **Genotype-by-Environment Interaction**. Boca Raton, Florida: CRC Press, 1996. chap. 9, p. 235-262.

HWANG, Y.T.; Wei, P.F. A novel method for testing normality in a mixed model of a nested classification. **Computational Statistics & Data Analysis**, Amsterdam, v. 51, n. 2, p. 1163-1183, Nov. 2006.

KENWARD, M.G.; ROGER, J.H. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. **Biometrics**, Arlington, v. 53, n. 3, p. 983-997, Sep. 1997.

LAIRD, N.M.; WARE, J.H. Random-Effects Models for longitudinal Data. **Biometrics**, Arlington, v. 38, n. 4, p. 963-974, Dec. 1982.

LANGE, N.; RYAN, L. Assessing normality in random effects models. **The Annals of Statistics**, California, v. 17, n. 2, p. 624-642, 1989.

LASAFFRE, E.; VERBEKE, G. Local Influence in Linear Mixed Models. **Biometrics**, Arlington, v. 54, n. 2, p. 570-582, Jun. 1998.

MACHADO, A. de A.; SILVA, J.G.C. da; DEMÉTRIO, C.G.B.; FERREIRA, D.F. Estatística experimental: Uma abordagem baseada no planejamento e no uso de recursos computacionais. In: Simpósio de Estatística Aplicada à Experimentação Agronômica, 11., 2005, Londrina. **Minicurso**. Londrina: UEL, 2005. 300 p.

MOLENBERGHS, G.; VERBEKE, G. **Models for Discrete Longitudinal Data**. New York: Springer-Verlag, 2005. 683 p.

NOBRE, J.S. **Métodos de diagnósticos para modelos lineares mistos**. 2004. 99 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2004.

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

PEREIRA, R.A.G. **Estudo comparativo das propriedades físico-mecânicas da celulose sulfato de madeira de *Eucalyptus saligna* Smith, *Eucalyptus alba* Reinw e *Eucalyptus grandis* Hill ex Maiden**. 1969. 129 p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1969.

PETERSEN, K.B.; PEDERSEN, M.S. **The Matrix Cookbook**. Disponível em: <<http://matrixcookbook>>. Acesso em 14 nov. 2008.

PIEPHO, H.P.; BÜCHSE, A.; EMRICH, K. A Hitchhiker's Guide to Mixed Models for Randomized Experiments. **Agronomy & Crop Science**, Berlin, v. 189, n. 5, p. 310-322, Oct. 2003.

PIEPHO, H.P.; BÜCHSE, A.; RICHTER, C. A Mixed Modelling Approach for Randomized Experiments with Repeated Measures. **Agronomy & Crop Science**, Berlin, v. 190, n. 4, p. 230-247, Aug. 2004.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528 p.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

RAUDENBUSH, S.W.; BRYK, A.S. **Hierarchical Linear Models: Applications and Data Analysis Methods**. 2 Edição. Thousand Oaks: Sage Publications, Inc., 2002. 457 p.

SATTERTHWAITE, F.E. An Approximate Distribution of Estimates of Variance Components. **Biometrics Bulletin**, Washington, v. 2, n. 6, p. 110-114, Dec. 1946.

SCHABENBERGER, O. Mixed Model Influence Diagnostics, in Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference, Paper 189-29, Cary, NC: SAS Institute, 2004.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Jerusalem, v. 6, n. 2, p. 461-464, 1978.

SEARLE, S.R.; CASELLA, G.; MCCULLOCH, C.E. **Variance components**. New York: John Wiley, 1992. p. 536.

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**, Alexandria, v. 82, n. 398, p. 605-610, 1987.

SMITH, A.; CULLIS, B.; GILMOUR, A. The analysis of crop variety evaluation data in Australia. **Australian and New Zealand Journal of Statistics**, Oxford, v. 43, n. 2, p. 129-145, 2001.

SMYTH, G.K.; VERBYLA, A.P. A Conditional Likelihood Approach to Residual Maximum Likelihood Estimation in Generalized Linear Models. **Journal of the Royal Statistical Society**, London, v. 58, n. 3, p. 565-572, 1996.

SNIJDERS, T.; BOSKER, R. **Multilevel Analysis: An introduction to basic and advanced multilevel modeling**. 1 Edição. Thousand Oaks: Sage Publications Inc., 1999. 251 p.

STRAM, D.O.; LEE, J.W. Variance Components Testing in the Longitudinal Mixed Effects Model. **Biometrics**, Arlington, v. 50, n. 4, p. 1171-1177, 1994.

TAYLOR Jr., W.H.; HILTON, H.G. A Structure Diagram Symbolization for Analysis of Variance. **The American Statistician**, New York, v. 35, n. 2, p. 85-93, May. 1981.

VERBEKE, G. LESAFFRE, E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. **Computacional Statistics & Data Analysis**, Amsterdam, v. 23, n. 4, p. 541-556, 1997.

VERBEKE, G.; MOLENBERGHS. **Linear mixed models for longitudinal data**. New York: Springer-Verlag, 2000. 568 p.

VERBYLA, A.P.; CULLIS, B.R.; KENWARD, M.G. WELHAN,S.J. The Analysis of Designed Experiments and Longitudinal Data by Using Smoothing Splines. **Journal of the Royal Statistical Society**, London, v. 48, n. 3, p. 269-311, 1999.

WALSH, B. **Mixed Models**. Disponível em: <<http://nitro.biosci.arizona.edu/courses/EEB596/handouts/mixed.pdf>>. Acesso em: 19 nov. 2010.

WEST, B.T.; WELCH, K.B.; GALECKI, A.T. **Linear mixed models: A pactical guide using statistical software**. New York: Chapman & Hall, 2007. 376 p.

WOLFINGER, R.D.; TOBIAS, R.D. Joint Estimation of Location, Dispersion, and Random Effects in Robust Design. **Technometrics**, Alexandria, v. 40, n. 1, p. 62-71, Feb. 1998.

ZEWOTIR, T.; GALPIN, J.S. Influence diagnostics for linear mixed models. **Journal of Data Science**, New York, v. 3, n. 2, p. 153-177. Apr. 2005.

ANEXOS

ANEXO A - Códigos para ajuste do modelo para os dados do experimento com instalações avícolas

```

require(car)
require(dae)
require(asreml)
require(sundry)

dados <- asreml.read.table("dados4.csv", header=T, dec=".", sep=";",
                           na.strings = ".")
dados$BP<- fac.combine(list(dados$Bloco, dados$Parc))
dados$Hora<- as.factor(dados$hora)

plot.profiles(formula=t ~ hora | Telha*Forro , groups.factor=BP,
               data=dados, key=F, xlab="Período (horas)",
               ylab="temperatura de bulbo seco (°C)", black=T, title=" ",
               scales=list(x=list(at=seq(1,4,1))))

X11()
par(mfrow=c(2,2))
###
TC_FA<- tapply(dados$t[dados$Telha=="C" & dados$Forro=="A"],
               dados$hora[dados$Telha=="C" & dados$Forro=="A"], mean)
plot(TC_FA, ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     main="Telhas:Forros : C:A", type="l", axes=F, col="red", lwd=2,
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
axis(1, at=seq(1,4,.5), labels =seq(8,17,1.5))
axis(2, at=seq(min(dados$t)-.005, max(dados$t)+.005, length=6),
     labels =seq(min(dados$t)-.005, max(dados$t)+.005, , length=6))
box()
for(i in 1:15){
  lines(dados$t[dados$Telha=="C" & dados$Forro=="A"][(4*i-3):(4*i)],, lty=2)
  points(dados$t[dados$Telha=="C" & dados$Forro=="A"][(4*i-3):(4*i)])
}
###
TC_FB<- tapply(dados$t[dados$Telha=="C" & dados$Forro=="B"],
               dados$hora[dados$Telha=="C" & dados$Forro=="B"], mean)
plot(TC_FB, ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     main="Telhas:Forros : C:B", type="l", axes=F, col="red", lwd=2,
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
axis(1, at=seq(1,4,.5), labels =seq(8,17,1.5))
axis(2, at=seq(min(dados$t)-.005, max(dados$t)+.005, length=6),
     labels =seq(min(dados$t)-.005, max(dados$t)+.005, , length=6))
box()
for(i in 1:15){
  lines(dados$t[dados$Telha=="C" & dados$Forro=="B"][(4*i-3):(4*i)],, lty=2)
  points(dados$t[dados$Telha=="C" & dados$Forro=="B"][(4*i-3):(4*i)])
}
###
TF_FA<- tapply(dados$t[dados$Telha=="F" & dados$Forro=="A"],
               dados$hora[dados$Telha=="F" & dados$Forro=="A"], mean)
plot(TF_FA, ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     main="Telhas:Forros : F:A", type="l", axes=F, col="red", lwd=2,

```

```

      ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
axis(1, at=seq(1,4,.5), labels =seq(8,17,1.5))
axis(2, at=seq(min(dados$t)-.005, max(dados$t)+.005, length=6),
      labels =seq(min(dados$t)-.005, max(dados$t)+.005, , length=6))
box()
for(i in 1:15){
lines(dados$t[dados$Telha=="F" & dados$Forro=="A"][(4*i-3):(4*i)],, lty=2)
points(dados$t[dados$Telha=="F" & dados$Forro=="A"][(4*i-3):(4*i)])
}
###
TF_FB<- tapply(dados$t[dados$Telha=="F" & dados$Forro=="B"],
               dados$hora[dados$Telha=="F" & dados$Forro=="B"], mean)
plot(TF_FB, ylim=c(min(dados$t)-.005, max(dados$t)+.005),
      main="Telhas:Forros : F:B", type="l", axes=F, col="red", lwd=2,
      ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
axis(1, at=seq(1,4,.5), labels =seq(8,17,1.5))
axis(2, at=seq(min(dados$t)-.005, max(dados$t)+.005, length=6),
      labels =seq(min(dados$t)-.005, max(dados$t)+.005, , length=6))
box()
for(i in 1:15){
lines(dados$t[dados$Telha=="F" & dados$Forro=="B"][(4*i-3):(4*i)],, lty=2)
points(dados$t[dados$Telha=="F" & dados$Forro=="B"][(4*i-3):(4*i)])
}

mod1<- asreml(t ~ Forro*Telha*Hora,
              random=~ Bloco/Hora + Bloco:Parc,
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
wald.asreml(mod1, denDF="algebraic", ssType="conditional")
summary(mod1)$varcomp
info.crit(mod1)
#graficos(mod1, dados, long=c("hora"), Long=c("Hora"), hora$t, dados$BP)

# modelo reparametrizado
mod2<- asreml(t ~ Telha*Forro*hora,
              random=~ spl(hora)/(Telha*Forro) +
                dev(hora)/(Telha*Forro) +
                str(~Bloco/hora, ~diag(2):id(15)) + Bloco:pol(hora,2) +
                str(~BP/hora, ~diag(2):id(60)) + BP:pol(hora,2) +
                Bloco:(spl(hora) + dev(hora))+
                BP:spl(hora),
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
mod2<- update(mod2)
summary(mod2)$varcomp
info.crit(mod1)
info.crit(mod2)
graficos(mod2, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# removendo os termos de fronteira
mod2.1<- asreml(t ~ Telha*Forro*hora,

```

```

        random=~ spl(hora):Forro/Telha +
                Bloco:(lin(Hora) + pol(hora,2)) +
                BP:(lin(Hora) + pol(hora,2)) +
                Bloco:(spl(hora) + dev(hora)),
        rcov=~ Bloco:Parc:Hora,
        aom=T,
        data=dados)
summary(mod2.1)$varcomp
info.crit(mod1)
info.crit(mod2)
info.crit(mod2.1)
reml.lrt(mod2, mod2.1, positive.zero=T) # fico com mod2.1
graficos(mod2.1, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - BP:pol(hora, 2)
mod3<- asreml(t ~ Telha*Forro*hora,
        random=~ spl(hora):Forro/Telha +
                Bloco:(lin(Hora) + pol(hora,2)) +
                BP:lin(Hora) +
                Bloco:(spl(hora) + dev(hora)),
        rcov=~ Bloco:Parc:Hora,
        aom=T,
        data=dados)
summary(mod3)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod3)
reml.lrt(mod2.1, mod3, positive.zero=T) # fico com mod2.1
graficos(mod3, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

#-spl(hora):Forro:Telha
mod4<- asreml(t ~ Telha*Forro*hora,
        random=~ spl(hora):Forro +
                Bloco:(lin(Hora) + pol(hora,2)) +
                Bloco:Parc:(lin(Hora) + pol(hora,2)) +
                Bloco:(spl(hora) + dev(hora)),
        rcov=~ Bloco:Parc:Hora,
        aom=T,
        data=dados)
summary(mod4)$varcomp
info.crit(mod1)
info.crit(mod2)
info.crit(mod4)
reml.lrt(mod2.1, mod4, positive.zero=T) # fico com mod2.1
graficos(mod4, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - lin(Hora):BP
mod5<- asreml(t ~ Telha*Forro*hora,
        random=~ spl(hora):Forro/Telha +
                Bloco:(lin(Hora) + pol(hora,2)) +
                Bloco:Parc:pol(hora,2) +
                Bloco:(spl(hora) + dev(hora)),
        rcov=~ Bloco:Parc:Hora,

```

```

        aom=T,
        data=dados)
summary(mod5)$varcomp
info.crit(mod1)
info.crit(mod2)
info.crit(mod5)
reml.lrt(mod2.1, mod5, positive.zero=T) # fico com mod5
graficos(mod5, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - Bloco:pol(hora, 2)
mod6<- asreml(t ~ Telha*Forro*hora,
              random=~ spl(hora):Forro/Telha +
                    Bloco:(lin(Hora)) +
                    Bloco:Parc:pol(hora,2) +
                    Bloco:(spl(hora) + dev(hora)),
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
summary(mod6)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod5)
info.crit(mod6)
reml.lrt(mod5, mod6, positive.zero=T) # fico com mod6
graficos(mod6, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - spl(hora):Bloco
mod7<- asreml(t ~ Telha*Forro*hora,
              random=~ spl(hora):Forro/Telha +
                    Bloco:(lin(Hora)) +
                    Bloco:Parc:pol(hora,2) +
                    Bloco:(dev(hora)),
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
summary(mod7)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod6)
info.crit(mod7)
reml.lrt(mod6, mod7, positive.zero=T) # fico com mod7
graficos(mod7, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - spl(hora):Forro
mod8<- asreml(t ~ Telha*Forro*hora,
              random=~ spl(hora):Forro:Telha +
                    Bloco:(lin(Hora)) +
                    Bloco:Parc:pol(hora,2) +
                    Bloco:(dev(hora)),
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
summary(mod8)$varcomp

```



```

info.crit(mod1)
info.crit(mod2.1)
info.crit(mod6)
info.crit(mod8)
reml.lrt(mod7, mod8, positive.zero=T) # fico com mod7
graficos(mod8, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# - spl(hora):Forro:Telha
mod9<- asreml(t ~ Telha*Forro*hora,
              random=~ spl(hora):Forro +
                        Bloco:(lin(Hora)) +
                        Bloco:Parc:pol(hora,2) +
                        Bloco:(dev(hora)),
              rcov=~ Bloco:Parc:Hora,
              aom=T,
              data=dados)
summary(mod9)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod6)
info.crit(mod9)
reml.lrt(mod7, mod9, positive.zero=T) # fico com mod7
graficos(mod9, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

# Bloco:Parc:pol(hora, 2)
mod10<- asreml(t ~ Telha*Forro*hora,
               random=~ spl(hora):Forro/Telha +
                        Bloco:(lin(Hora)) +
                        Bloco:(dev(hora)),
               rcov=~ Bloco:Parc:Hora,
               aom=T,
               data=dados)
summary(mod10)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod6)
info.crit(mod10)
reml.lrt(mod7, mod10, positive.zero=T) # fico com mod7
graficos(mod10, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

#modificando R
mod7.ar1<- asreml(t ~ Telha*Forro*hora,
                  random=~ spl(hora):Forro/Telha +
                            Bloco:(lin(Hora)) +
                            Bloco:Parc:pol(hora,2) +
                            Bloco:(dev(hora)),
                  rcov=~ Bloco:Parc:ar1(Hora),
                  aom=T,
                  data=dados)
summary(mod7.ar1)$varcomp
info.crit(mod1)
info.crit(mod2.1)
info.crit(mod7)

```

```

info.crit(mod7.ar1)
reml.lrt(mod7.ar1, mod7, positive.zero=F) # fico com mod7.ar1
graficos(mod7.ar1, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

mod7.idh<- asreml(t ~ Telha*Forro*hora,
  random=~ spl(hora):Forro/Telha +
    Bloco:(lin(Hora)) +
    Bloco:Parc:pol(hora,2) +
    Bloco:(dev(hora)),
  rcov=~ Bloco:Parc:idh(Hora),
  aom=T,
  data=dados)
summary(mod7.idh)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod7.ar1)
info.crit(mod7.idh)
reml.lrt(mod7.idh, mod7.ar1, positive.zero=F) # fico com mod7.idh
graficos(mod7.idh, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

mod7.ar1h<- asreml(t ~ Telha*Forro*hora,
  random=~ spl(hora):Forro/Telha +
    Bloco:(lin(Hora)) +
    Bloco:Parc:pol(hora,2) +
    Bloco:(dev(hora)),
  rcov=~ Bloco:Parc:ar1h(Hora),
  aom=T,
  data=dados)
summary(mod7.ar1h)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod7.ar1)
info.crit(mod7.idh)
info.crit(mod7.ar1h)
reml.lrt(mod7.ar1h, mod7.idh, positive.zero=F) # fico com mod7.idh
graficos(mod7.ar1h, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

mod7.corgh<- asreml(t ~ Telha*Forro*hora,
  random=~ spl(hora):Forro/Telha +
    Bloco:(lin(Hora)) +
    Bloco:(dev(hora)),
  rcov=~ Bloco:Parc:corgh(Hora),
  aom=T,
  data=dados)
summary(mod7.corgh)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod7.ar1)
info.crit(mod7.idh)
info.crit(mod7.corgh)
reml.lrt(mod7.corgh, mod7.idh, positive.zero=F) # fico com mod7.idh
graficos(mod7.corgh, dados, long=c("hora"), Long=c("Hora"), dados$t, dados$BP)

```

```

# supondo que eu fique com o modelo mod7.idh
# teste wald-F
wald.asreml(mod7.idh, denDF="algebraic", ssType="conditional")

mod7.idh.fix<- asreml(t ~ Telha*hora + Forro + Telha:Forro:hora,
  random=~ spl(hora):Forro/Telha +
    Bloco:(lin(Hora)) +
    Bloco:Parc:pol(hora,2) +
    Bloco:(dev(hora)),
  rcov=~ Bloco:Parc:idh(Hora),
  aom=T,
  data=dados)
wald.asreml(mod7.idh.fix, denDF="algebraic", ssType="conditional")

mod7.idh.fix2<- asreml(t ~ Telha*hora + Forro,
  random=~ spl(hora):Forro/Telha +
    Bloco:(lin(Hora)) +
    Bloco:Parc:pol(hora,2) +
    Bloco:(dev(hora)),
  rcov=~ Bloco:Parc:idh(Hora),
  aom=T,
  data=dados)
wald.asreml(mod7.idh.fix2, denDF="algebraic", ssType="conditional")
summary(mod7.idh.fix2)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod7.idh)
info.crit(mod7.idh.fix2)
graficos(mod7.idh.fix2, dados, long=c("hora"), Long=c("Hora"), dados$t,
  dados$BP)

mod7.idh.fix2.2<- asreml(t ~ Telha*hora + Forro,
  random=~ spl(hora) +
    Bloco:lin(Hora) +
    Bloco:pol(hora,2) +
    Bloco:dev(hora),
  rcov=~ Bloco:Parc:idh(Hora),
  aom=T,
  data=dados)
wald.asreml(mod7.idh.fix2.2, denDF="algebraic", ssType="conditional")
summary(mod7.idh.fix2.2)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod7.idh.fix2)
info.crit(mod7.idh.fix2.2)
graficos(mod7.idh.fix2.2, dados, long=c("hora"), Long=c("Hora"), dados$t,
  dados$BP)
reml.lrt(mod7.idh.fix2, mod7.idh.fix2.2, positive.zero=F) # fico com mod7.idh

pv <- predict(mod7.idh.fix2.2, classify = "Telha:Forro:hora",
  predictpoints=list(tempo=seq(8,17,3)))
pred <- pv$predictions$pvals
pred$Telha <- factor(pred$Telha)

```

```

pred$Forro <- factor(pred$Forro)

X11()
par(mfrow=c(2,2))
plot(t ~ hora, data=dados[dados$Telha=="C" & dados$Forro=="A",],
     main="Telhas:Forros : C:A", ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="A",],
     col=2, lwd=2)
#
plot(t ~ hora, data=dados[dados$Telha=="C" & dados$Forro=="B",],
     main="Telhas:Forros : C:B", ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="B",],
     col=2, lwd=2)
#
plot(t ~ hora, data=dados[dados$Telha=="F" & dados$Forro=="A",],
     main="Telhas:Forros : F:A", ylim=c(min(dados$t)-.005, max(dados$t)+.005),
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="A",],
     col=2, lwd=2)
#
plot(t ~ hora, data=dados[dados$Telha=="F" & dados$Forro=="B",],
     main="Telhas:Forros : F:B", ylim=c(min(dados$t)-.005,
     max(dados$t)+.005), ylab="temperatura de bulbo seco (°C)",
     xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="B",],
     col=2, lwd=2)

X11()
plot(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="A",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=1, type="l",
     ylab="temperatura de bulbo seco (°C)", xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="B",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=2)
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="A",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=3)
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="B",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=4)
legend(13.5, 33.5, c("Telhas:Forros : C:A", "Telhas:Forros : C:B",
                    "Telhas:Forros : F:A", "Telhas:Forros : F:B"), lty=c(1:4))

X11()
plot(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="A",],
     lty=1, type="l", ylab="temperatura de bulbo seco (°C)",
     xlab="tempo (horas)")
lines(predicted.value ~ hora, data=pred[pred$Telha=="C" & pred$Forro=="B",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=2)
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="A",],
     ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=3)

```

```
lines(predicted.value ~ hora, data=pred[pred$Telha=="F" & pred$Forro=="B",],  
      ylim=c(min(dados$t)-.005, max(dados$t)+.005), lty=4)  
legend(13.5, 33.5, c("Telhas:Forros : C:A", "Telhas:Forros : C:B",  
                    "Telhas:Forros : F:A", "Telhas:Forros : F:B"), lty=c(1:4))
```

ANEXO B - Dados de peso específico aparente (g/cm^3) do experimento com polpas de Eucalipto

Tabela 16 - Dados do experimento com polpas de Eucalipto

(continua)							
Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
37	1	E. alba	5	1	1	30	0,599
37	2	E. alba	5	1	1	60	0,642
37	3	E. alba	5	1	1	90	0,646
37	4	E. alba	5	1	1	120	0,672
37	5	E. alba	5	1	1	150	0,698
37	6	E. alba	5	1	1	180	0,737
38	1	E. alba	5	1	2	30	0,620
38	2	E. alba	5	1	2	60	0,643
38	3	E. alba	5	1	2	90	0,677
38	4	E. alba	5	1	2	120	0,675
38	5	E. alba	5	1	2	150	0,728
38	6	E. alba	5	1	2	180	0,773
39	1	E. alba	5	1	3	30	0,624
39	2	E. alba	5	1	3	60	0,638
39	3	E. alba	5	1	3	90	0,661
39	4	E. alba	5	1	3	120	0,684
39	5	E. alba	5	1	3	150	0,673
39	6	E. alba	5	1	3	180	0,744
40	1	E. alba	5	1	4	30	0,628
40	2	E. alba	5	1	4	60	0,659
40	3	E. alba	5	1	4	90	0,687
40	4	E. alba	5	1	4	120	0,706
40	5	E. alba	5	1	4	150	0,735
40	6	E. alba	5	1	4	180	0,765
41	1	E. alba	5	2	1	30	0,587
41	2	E. alba	5	2	1	60	0,610
41	3	E. alba	5	2	1	90	0,634
41	4	E. alba	5	2	1	120	0,632
41	5	E. alba	5	2	1	150	0,668
41	6	E. alba	5	2	1	180	0,699
42	1	E. alba	5	2	2	30	0,575
42	2	E. alba	5	2	2	60	0,607
42	3	E. alba	5	2	2	90	0,643
42	4	E. alba	5	2	2	120	0,655
42	5	E. alba	5	2	2	150	0,684
42	6	E. alba	5	2	2	180	0,747
43	1	E. alba	5	2	3	30	0,585
43	2	E. alba	5	2	3	60	0,610
43	3	E. alba	5	2	3	90	0,628
43	4	E. alba	5	2	3	120	0,645
43	5	E. alba	5	2	3	150	0,685
43	6	E. alba	5	2	3	180	0,696
44	1	E. alba	5	2	4	30	0,579
44	2	E. alba	5	2	4	60	0,600
44	3	E. alba	5	2	4	90	0,629

Tabela 16 - Dados do experimento com polpas de Eucalipto

(continuação)

Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
44	4	E. alba	5	2	4	120	0,649
44	5	E. alba	5	2	4	150	0,689
44	6	E. alba	5	2	4	180	0,742
45	1	E. alba	5	3	1	30	0,606
45	2	E. alba	5	3	1	60	0,630
45	3	E. alba	5	3	1	90	0,650
45	4	E. alba	5	3	1	120	0,658
45	5	E. alba	5	3	1	150	0,675
45	6	E. alba	5	3	1	180	0,691
46	1	E. alba	5	3	2	30	0,620
46	2	E. alba	5	3	2	60	0,643
46	3	E. alba	5	3	2	90	0,670
46	4	E. alba	5	3	2	120	0,687
46	5	E. alba	5	3	2	150	0,717
46	6	E. alba	5	3	2	180	0,742
47	1	E. alba	5	3	3	30	0,609
47	2	E. alba	5	3	3	60	0,632
47	3	E. alba	5	3	3	90	0,662
47	4	E. alba	5	3	3	120	0,685
47	5	E. alba	5	3	3	150	0,703
47	6	E. alba	5	3	3	180	0,771
48	1	E. alba	5	3	4	30	0,603
48	2	E. alba	5	3	4	60	0,634
48	3	E. alba	5	3	4	90	0,662
48	4	E. alba	5	3	4	120	0,668
48	5	E. alba	5	3	4	150	0,704
48	6	E. alba	5	3	4	180	0,757
25	1	E. alba	7	1	1	30	0,577
25	2	E. alba	7	1	1	60	0,602
25	3	E. alba	7	1	1	90	0,625
25	4	E. alba	7	1	1	120	0,632
25	5	E. alba	7	1	1	150	0,646
25	6	E. alba	7	1	1	180	0,685
26	1	E. alba	7	1	2	30	0,593
26	2	E. alba	7	1	2	60	0,626
26	3	E. alba	7	1	2	90	0,625
26	4	E. alba	7	1	2	120	0,646
26	5	E. alba	7	1	2	150	0,664
26	6	E. alba	7	1	2	180	0,685
27	1	E. alba	7	1	3	30	0,590
27	2	E. alba	7	1	3	60	0,606
27	3	E. alba	7	1	3	90	0,641
27	4	E. alba	7	1	3	120	0,640
27	5	E. alba	7	1	3	150	0,667
27	6	E. alba	7	1	3	180	0,684
28	1	E. alba	7	1	4	30	0,586
28	2	E. alba	7	1	4	60	0,602
28	3	E. alba	7	1	4	90	0,617
28	4	E. alba	7	1	4	120	0,628

Tabela 16 - Dados do experimento com polpas de Eucalipto

(continuação)

Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
28	5	E. alba	7	1	4	150	0,656
28	6	E. alba	7	1	4	180	0,687
29	1	E. alba	7	2	1	30	0,612
29	2	E. alba	7	2	1	60	0,634
29	3	E. alba	7	2	1	90	0,655
29	4	E. alba	7	2	1	120	0,663
29	5	E. alba	7	2	1	150	0,632
29	6	E. alba	7	2	1	180	0,732
30	1	E. alba	7	2	2	30	0,630
30	2	E. alba	7	2	2	60	0,639
30	3	E. alba	7	2	2	90	0,656
30	4	E. alba	7	2	2	120	0,700
30	5	E. alba	7	2	2	150	0,727
30	6	E. alba	7	2	2	180	0,729
31	1	E. alba	7	2	3	30	0,602
31	2	E. alba	7	2	3	60	0,642
31	3	E. alba	7	2	3	90	0,648
31	4	E. alba	7	2	3	120	0,697
31	5	E. alba	7	2	3	150	0,727
31	6	E. alba	7	2	3	180	0,737
32	1	E. alba	7	2	4	30	0,600
32	2	E. alba	7	2	4	60	0,627
32	3	E. alba	7	2	4	90	0,653
32	4	E. alba	7	2	4	120	0,699
32	5	E. alba	7	2	4	150	0,694
32	6	E. alba	7	2	4	180	0,730
33	1	E. alba	7	3	1	30	0,574
33	2	E. alba	7	3	1	60	0,589
33	3	E. alba	7	3	1	90	0,609
33	4	E. alba	7	3	1	120	0,634
33	5	E. alba	7	3	1	150	0,663
33	6	E. alba	7	3	1	180	0,710
34	1	E. alba	7	3	2	30	0,571
34	2	E. alba	7	3	2	60	0,587
34	3	E. alba	7	3	2	90	0,599
34	4	E. alba	7	3	2	120	0,621
34	5	E. alba	7	3	2	150	0,634
34	6	E. alba	7	3	2	180	0,676
35	1	E. alba	7	3	3	30	0,560
35	2	E. alba	7	3	3	60	0,586
35	3	E. alba	7	3	3	90	0,620
35	4	E. alba	7	3	3	120	0,636
35	5	E. alba	7	3	3	150	0,673
35	6	E. alba	7	3	3	180	0,704
36	1	E. alba	7	3	4	30	0,594
36	2	E. alba	7	3	4	60	0,606
36	3	E. alba	7	3	4	90	0,640
36	4	E. alba	7	3	4	120	0,650
36	5	E. alba	7	3	4	150	0,677

Tabela 16 - Dados do experimento com polpas de Eucalipto

(continuação)

Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
36	6	E. alba	7	3	4	180	0,714
1	1	E. saligna	5	1	1	30	0,623
1	2	E. saligna	5	1	1	60	0,643
1	3	E. saligna	5	1	1	90	0,657
1	4	E. saligna	5	1	1	120	0,689
1	5	E. saligna	5	1	1	150	0,733
1	6	E. saligna	5	1	1	180	0,749
2	1	E. saligna	5	1	2	30	0,633
2	2	E. saligna	5	1	2	60	0,651
2	3	E. saligna	5	1	2	90	0,656
2	4	E. saligna	5	1	2	120	0,692
2	5	E. saligna	5	1	2	150	0,718
2	6	E. saligna	5	1	2	180	0,735
3	1	E. saligna	5	1	3	30	0,620
3	2	E. saligna	5	1	3	60	0,648
3	3	E. saligna	5	1	3	90	0,689
3	4	E. saligna	5	1	3	120	0,686
3	5	E. saligna	5	1	3	150	0,716
3	6	E. saligna	5	1	3	180	0,748
4	1	E. saligna	5	1	4	30	0,621
4	2	E. saligna	5	1	4	60	0,644
4	3	E. saligna	5	1	4	90	0,670
4	4	E. saligna	5	1	4	120	0,673
4	5	E. saligna	5	1	4	150	0,699
4	6	E. saligna	5	1	4	180	0,744
5	1	E. saligna	5	2	1	30	0,674
5	2	E. saligna	5	2	1	60	0,696
5	3	E. saligna	5	2	1	90	0,728
5	4	E. saligna	5	2	1	120	0,744
5	5	E. saligna	5	2	1	150	0,780
5	6	E. saligna	5	2	1	180	0,865
6	1	E. saligna	5	2	2	30	0,632
6	2	E. saligna	5	2	2	60	0,667
6	3	E. saligna	5	2	2	90	0,679
6	4	E. saligna	5	2	2	120	0,706
6	5	E. saligna	5	2	2	150	0,729
6	6	E. saligna	5	2	2	180	0,756
7	1	E. saligna	5	2	3	30	0,622
7	2	E. saligna	5	2	3	60	0,654
7	3	E. saligna	5	2	3	90	0,696
7	4	E. saligna	5	2	3	120	0,727
7	5	E. saligna	5	2	3	150	0,745
7	6	E. saligna	5	2	3	180	0,817
8	1	E. saligna	5	2	4	30	0,643
8	2	E. saligna	5	2	4	60	0,666
8	3	E. saligna	5	2	4	90	0,697
8	4	E. saligna	5	2	4	120	0,727
8	5	E. saligna	5	2	4	150	0,776
8	6	E. saligna	5	2	4	180	0,862

Tabela 16 - Dados do experimento com polpas de Eucalipto

(continuação)

Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
9	1	E. saligna	5	3	1	30	0,631
9	2	E. saligna	5	3	1	60	0,670
9	3	E. saligna	5	3	1	90	0,697
9	4	E. saligna	5	3	1	120	0,718
9	5	E. saligna	5	3	1	150	0,765
9	6	E. saligna	5	3	1	180	0,774
10	1	E. saligna	5	3	2	30	0,611
10	2	E. saligna	5	3	2	60	0,637
10	3	E. saligna	5	3	2	90	0,669
10	4	E. saligna	5	3	2	120	0,680
10	5	E. saligna	5	3	2	150	0,712
10	6	E. saligna	5	3	2	180	0,749
11	1	E. saligna	5	3	3	30	0,643
11	2	E. saligna	5	3	3	60	0,674
11	3	E. saligna	5	3	3	90	0,674
11	4	E. saligna	5	3	3	120	0,725
11	5	E. saligna	5	3	3	150	0,800
11	6	E. saligna	5	3	3	180	0,830
12	1	E. saligna	5	3	4	30	0,620
12	2	E. saligna	5	3	4	60	0,656
12	3	E. saligna	5	3	4	90	0,694
12	4	E. saligna	5	3	4	120	0,716
12	5	E. saligna	5	3	4	150	0,755
12	6	E. saligna	5	3	4	180	0,839
13	1	E. saligna	7	1	1	30	0,618
13	2	E. saligna	7	1	1	60	0,651
13	3	E. saligna	7	1	1	90	0,652
13	4	E. saligna	7	1	1	120	0,678
13	5	E. saligna	7	1	1	150	0,707
13	6	E. saligna	7	1	1	180	0,742
14	1	E. saligna	7	1	2	30	0,604
14	2	E. saligna	7	1	2	60	0,630
14	3	E. saligna	7	1	2	90	0,636
14	4	E. saligna	7	1	2	120	0,657
14	5	E. saligna	7	1	2	150	0,680
14	6	E. saligna	7	1	2	180	0,702
15	1	E. saligna	7	1	3	30	0,622
15	2	E. saligna	7	1	3	60	0,651
15	3	E. saligna	7	1	3	90	0,651
15	4	E. saligna	7	1	3	120	0,668
15	5	E. saligna	7	1	3	150	0,679
15	6	E. saligna	7	1	3	180	0,701
16	1	E. saligna	7	1	4	30	0,592
16	2	E. saligna	7	1	4	60	0,636
16	3	E. saligna	7	1	4	90	0,645
16	4	E. saligna	7	1	4	120	0,679
16	5	E. saligna	7	1	4	150	0,703
16	6	E. saligna	7	1	4	180	0,740

Tabela 16 - Dados do experimento com polpas de Eucalipto

(conclusão)

Rodadas	Posições	Espécies	Idades	Lotes	Grupos	Tempos	peso
17	1	E. saligna	7	2	1	30	0,626
17	2	E. saligna	7	2	1	60	0,672
17	3	E. saligna	7	2	1	90	0,654
17	4	E. saligna	7	2	1	120	0,712
17	5	E. saligna	7	2	1	150	0,736
17	6	E. saligna	7	2	1	180	0,748
18	1	E. saligna	7	2	2	30	0,619
18	2	E. saligna	7	2	2	60	0,662
18	3	E. saligna	7	2	2	90	0,682
18	4	E. saligna	7	2	2	120	0,702
18	5	E. saligna	7	2	2	150	0,696
18	6	E. saligna	7	2	2	180	0,732
19	1	E. saligna	7	2	3	30	0,631
19	2	E. saligna	7	2	3	60	0,652
19	3	E. saligna	7	2	3	90	0,666
19	4	E. saligna	7	2	3	120	0,682
19	5	E. saligna	7	2	3	150	0,708
19	6	E. saligna	7	2	3	180	0,725
20	1	E. saligna	7	2	4	30	0,615
20	2	E. saligna	7	2	4	60	0,633
20	3	E. saligna	7	2	4	90	0,656
20	4	E. saligna	7	2	4	120	0,685
20	5	E. saligna	7	2	4	150	0,707
20	6	E. saligna	7	2	4	180	0,751
21	1	E. saligna	7	3	1	30	0,594
21	2	E. saligna	7	3	1	60	0,636
21	3	E. saligna	7	3	1	90	0,636
21	4	E. saligna	7	3	1	120	0,681
21	5	E. saligna	7	3	1	150	0,708
21	6	E. saligna	7	3	1	180	0,725
22	1	E. saligna	7	3	2	30	0,592
22	2	E. saligna	7	3	2	60	0,606
22	3	E. saligna	7	3	2	90	0,653
22	4	E. saligna	7	3	2	120	0,684
22	5	E. saligna	7	3	2	150	0,702
22	6	E. saligna	7	3	2	180	0,728
23	1	E. saligna	7	3	3	30	0,599
23	2	E. saligna	7	3	3	60	0,622
23	3	E. saligna	7	3	3	90	0,652
23	4	E. saligna	7	3	3	120	0,680
23	5	E. saligna	7	3	3	150	0,709
23	6	E. saligna	7	3	3	180	0,762
24	1	E. saligna	7	3	4	30	0,603
24	2	E. saligna	7	3	4	60	0,614
24	3	E. saligna	7	3	4	90	0,651
24	4	E. saligna	7	3	4	120	0,674
24	5	E. saligna	7	3	4	150	0,700
24	6	E. saligna	7	3	4	180	0,734

Fonte: Pereira (1969)

ANEXO C - Códigos para ajuste do modelo para os dados do experimento com polpas de Eucalipto

```

require(car)
require(dae)
require(asreml)
require(sundry)
#asreml.man()

Euc.Pulp2x2.dat <- asreml.read.table("peso.csv", header=T, dec=",", sep=";")
Euc.Pulp2x2.dat$EI <- fac.combine(list(Euc.Pulp2x2.dat$Especies,
                                      Euc.Pulp2x2.dat$Idades))
Euc.Pulp2x2.dat <- Euc.Pulp2x2.dat[ order(Euc.Pulp2x2.dat$Especies,
                                      Euc.Pulp2x2.dat$Idades,
                                      Euc.Pulp2x2.dat$Lotes,
                                      Euc.Pulp2x2.dat$Grupos,
                                      Euc.Pulp2x2.dat$Tempos), ]

Euc.Pulp2x2.dat$EIL <- factor(rep(1:12, each=24))
Euc.Pulp2x2.dat$EILG <- factor(rep(1:48, each=6))
Euc.Pulp2x2.dat$Lotes.Grupos <- fac.combine(list(Euc.Pulp2x2.dat$Lotes,
                                                  Euc.Pulp2x2.dat$Grupos))

save(Euc.Pulp2x2.dat, file="Euc.Pulp2x2_2.dat.rda")
load("Euc.Pulp2x2_2.dat.rda")
Euc.Pulp2x2.dat

# gráfico de perfis
plot.profiles(formula=peso ~ tempo | Idades:Especies,
              groups.factor=Lotes.Grupos,
              data=Euc.Pulp2x2.dat, key=F, xlab="tempo de refinação (min)",
              ylab="Peso específico aparente (g/cm3)", black=T, title=" ",
              scales=list(x=list(at=seq(30, 180, 60))))

X11()
par(mfrow=c(2,2))
###
I5_EA<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
                                   Euc.Pulp2x2.dat$Especies=="E. alba"],
              Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Idades==5 &
                                   Euc.Pulp2x2.dat$Especies=="E. alba"], mean)
plot(I5_EA, ylim=c(min(Euc.Pulp2x2.dat$peso)-.005,
                    max(Euc.Pulp2x2.dat$peso)+.005),
      xlab="tempo de refinação (min)", ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. alba:5", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(Euc.Pulp2x2.dat$peso)-.005,
                max(Euc.Pulp2x2.dat$peso)+.005, length=8),
      labels =seq(min(Euc.Pulp2x2.dat$peso)-.005,
                  max(Euc.Pulp2x2.dat$peso)+.005,
                  length=8))
box()
for(i in 1:12){

```

```

lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. alba"][(6*i-5):(6*i)],
      lty=2)
points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. alba"][(6*i-5):(6*i)])
}
###
I5_ES<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. saligna"],
      Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. saligna"],
      mean)
plot(I5_ES, ylim=c(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005),
      xlab="tempo de refinação (min)", ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. saligna:5", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005, length=8),
      labels =seq(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005, , length=8))
box()
for(i in 1:12){
lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. saligna"][(6*i-5):(6*i)],
      lty=2)
points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==5 &
      Euc.Pulp2x2.dat$Especies=="E. saligna"][(6*i-5):(6*i)])
}
###
I7_EA<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &
      Euc.Pulp2x2.dat$Especies=="E. alba"],
      Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Idades==7 &
      Euc.Pulp2x2.dat$Especies=="E. alba"], mean)
plot(I7_EA, ylim=c(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005),
      xlab="tempo de refinação (min)", ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. alba:7", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005, length=8),
      labels =seq(min(Euc.Pulp2x2.dat$peso)-.005,
      max(Euc.Pulp2x2.dat$peso)+.005, length=8))
box()
for(i in 1:12){
lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &
      Euc.Pulp2x2.dat$Especies=="E. alba"][(6*i-5):(6*i)],
      lty=2)
points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &
      Euc.Pulp2x2.dat$Especies=="E. alba"][(6*i-5):(6*i)])
}
###
I7_ES<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &

```

```

                                Euc.Pulp2x2.dat$Especies=="E. saligna"],
                                Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Idades==7 &
                                Euc.Pulp2x2.dat$Especies=="E. saligna"],
                                mean)
plot(I7_ES, ylim=c(min(Euc.Pulp2x2.dat$peso)-.005,
                      max(Euc.Pulp2x2.dat$peso)+.005),
      xlab="tempo de refinação (min)", ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. saligna:7", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(Euc.Pulp2x2.dat$peso)-.005,
                max(Euc.Pulp2x2.dat$peso)+.005, length=8),
      labels =seq(min(Euc.Pulp2x2.dat$peso)-.005,
                  max(Euc.Pulp2x2.dat$peso)+.005, length=8))
box()
for(i in 1:12){
  lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &
                             Euc.Pulp2x2.dat$Especies=="E. saligna"][(6*i-5):(6*i)],
        lty=2)
  points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Idades==7 &
                              Euc.Pulp2x2.dat$Especies=="E. saligna"][(6*i-5):(6*i)])
}

# modelo base
mod1 <- asreml(peso ~ Especies*Idades*Tempos,
               random= ~ Especies:Idades:Lotes/Tempos +
                       Especies:Idades:Lotes:Grupos,
               rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
               aom=T,
               data=Euc.Pulp2x2.dat)

summary(mod1)$varcomp
info.crit(mod1)
#graficos(mod1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
#          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

# modelo reparametrizado geral
mod2<- asreml(peso ~ Especies*Idades*tempo,
              random=~ spl(tempo)/(Especies*Idades) +
                      dev(tempo)/(Especies*Idades) +
                      str(~ EIL/tempo, ~ diag(2):id(12)) +
                      str(~ EILG/tempo, ~ diag(2):id(48)) +
                      EIL:(Tempos + spl(tempo)) +
                      EILG:(spl(tempo)),
              rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod2)$varcomp
info.crit(mod2)
graficos(mod2, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

#removendo termos de fronteira
mod2.1<- asreml(peso ~ Especies*Idades*tempo,

```

```

        random=~ spl(tempo)/Idades +
                dev(tempo):Idades +
                str(~ EIL/tempo, ~ diag(2):id(12)) +
                str(~ EILG/tempo, ~ diag(2):id(48)) +
                EILG:(spl(tempo)),
        rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
        aom=T,
        data= Euc.Pulp2x2.dat)

summary(mod2.1)$varcomp
info.crit(mod2.1)
graficos(mod2.1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
        Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod2, mod2.1, positive.zero=T) # fico com mod2.1

# -EILG int-coef + EILG componente
mod3<- asreml(peso ~ Especies*Idades*tempo,
        random=~ spl(tempo)/Idades +
                dev(tempo):Idades +
                str(~ EIL/tempo, ~ diag(2):id(12)) +
                EILG +
                EILG:(spl(tempo)),
        rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
        aom=T,
        data= Euc.Pulp2x2.dat)

summary(mod3)$varcomp
info.crit(mod2.1)
info.crit(mod3)
reml.lrt(mod2.1, mod3, positive.zero=T) # fico com mod2.1
graficos(mod3, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
        Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

# -spl(tempo):EILG
mod4<- asreml(peso ~ Especies*Idades*tempo,
        random=~ spl(tempo)/Idades +
                dev(tempo):Idades +
                str(~ EIL/tempo, ~ diag(2):id(12)) +
                str(~ EILG/tempo, ~ diag(2):id(48)),
        rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
        aom=T,
        data= Euc.Pulp2x2.dat)

summary(mod4)$varcomp
info.crit(mod2.1)
info.crit(mod4)
graficos(mod4, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
        Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod2.1, mod4, positive.zero=T) # fico com mod4

# -spl(tempo):Idades
mod5<- asreml(peso ~ Especies*Idades*tempo,
        random=~ spl(tempo) +
                dev(tempo):Idades +
                str(~ EIL/tempo, ~ diag(2):id(12)) +
                str(~ EILG/tempo, ~ diag(2):id(48)),

```

```

rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
aom=T,
data= Euc.Pulp2x2.dat)

summary(mod5)$varcomp
info.crit(mod2.1)
info.crit(mod4)
info.crit(mod5)
graficos(mod5, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod4, mod5, positive.zero=T) # fico com mod5

# -spl(tempo)
mod6<- asreml(peso ~ Especies*Idades*tempo,
              random=~ dev(tempo):Idades +
                    str(~ EIL/tempo, ~ diag(2):id(12)) +
                    str(~ EILG/tempo, ~ diag(2):id(48)),
              rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod6)$varcomp
info.crit(mod2.1)
info.crit(mod4)
info.crit(mod5)
info.crit(mod6)
graficos(mod6, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod5, mod6, positive.zero=T) # fico com mod5

# -dev(tempo):Idades
mod7<- asreml(peso ~ Especies*Idades*tempo,
              random=~ spl(tempo) +
                    str(~ EIL/tempo, ~ diag(2):id(12)) +
                    str(~ EILG/tempo, ~ diag(2):id(48)),
              rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod7)$varcomp
info.crit(mod2.1)
info.crit(mod5)
info.crit(mod7)
graficos(mod7, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod5, mod7, positive.zero=T) # fico com mod5

# -EILG int-coef + EILG componente
mod8<- asreml(peso ~ Especies*Idades*tempo,
              random=~ spl(tempo) +
                    dev(tempo):Idades +
                    str(~ EIL/tempo, ~ diag(2):id(12)) +
                    EILG,
              rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

```



```

summary(mod8)$varcomp
info.crit(mod2.1)
info.crit(mod5)
info.crit(mod8)
graficos(mod8, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod5, mod8, positive.zero=T) # fico com mod5

# -EIL int-coef + EIL componente
mod9<- asreml(peso ~ Especies*Idades*tempo,
              random=~ spl(tempo) +
                  dev(tempo):Idades +
                  EIL +
                  str(~ EILG/tempo, ~ diag(2):id(48)),
              rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod9)$varcomp
info.crit(mod2.1)
info.crit(mod5)
info.crit(mod9)
graficos(mod9, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod5, mod9, positive.zero=T) # fico com mod5

# + EILG cov t
mod10<- asreml(peso ~ Especies*Idades*tempo,
               random=~ spl(tempo) +
                   dev(tempo):Idades +
                   str(~ EIL/tempo, ~ diag(2):id(12)) +
                   str(~ EILG/tempo, ~ us(2):id(48)),
               rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
               aom=T,
               data= Euc.Pulp2x2.dat)

summary(mod10)$varcomp
info.crit(mod2.1)
info.crit(mod5)
info.crit(mod10)
graficos(mod10, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10, mod5, positive.zero=F) # fico com mod10

# - dev(tempo):Idades
mod11<- asreml(peso ~ Especies*Idades*tempo,
               random=~ spl(tempo) +
                   str(~ EIL/tempo, ~ diag(2):id(12)) +
                   str(~ EILG/tempo, ~ us(2):id(48)),
               rcov=~ Especies:Idades:Lotes:Grupos:Tempos,
               aom=T,
               data= Euc.Pulp2x2.dat)

summary(mod11)$varcomp
info.crit(mod2.1)
info.crit(mod10)

```

```

info.crit(mod11)
graficos(mod11, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10, mod11, positive.zero=T) # fico com mod10

# modificando Sigma
mod10.ar1<- asreml(peso ~ Especies*Idades*tempo,
                  random=~ spl(tempo) +
                        dev(tempo):Idades +
                        str(~ EIL/tempo, ~ diag(2):id(12)) +
                        str(~ EILG/tempo, ~ us(2):id(48)),
                  rcov=~ Especies:Idades:Lotes:Grupos:ar1(Tempos),
                  aom=T,
                  data= Euc.Pulp2x2.dat)
summary(mod10.ar1)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.ar1)
graficos(mod10.ar1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.ar1, mod10, positive.zero=F) # fico com mod10

#trocar us por diag em EILG:lin(tempo) para evitar dependência
mod10.idh<- asreml(peso ~ Especies*Idades*tempo,
                  random=~ spl(tempo) +
                        dev(tempo):Idades +
                        str(~ EIL/tempo, ~ diag(2):id(12)) +
                        str(~ EILG/tempo, ~ diag(2):id(48)),
                  rcov=~ Especies:Idades:Lotes:Grupos:indh(Tempos),
                  aom=T,
                  data= Euc.Pulp2x2.dat)
summary(mod10.idh)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.idh)
graficos(mod10.idh, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh, mod10, positive.zero=F) # fico com mod10.idh

mod10.ar1h<- asreml(peso ~ Especies*Idades*tempo,
                  random=~ spl(tempo) +
                        dev(tempo):Idades +
                        str(~ EIL/tempo, ~ diag(2):id(12)) +
                        str(~ EILG/tempo, ~ diag(2):id(48)),
                  rcov=~ Especies:Idades:Lotes:Grupos:ar1h(Tempos),
                  aom=T,
                  data= Euc.Pulp2x2.dat)
summary(mod10.ar1h)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.idh)
info.crit(mod10.ar1h)

```

```

graficos(mod10.ar1h, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.ar1h, mod10.idh, positive.zero=F) # fico com mod10.idh

gamm<- asreml(peso ~ Especies*Idades*tempo,
              random=~ spl(tempo) +
                    dev(tempo):Idades +
                    str(~ EIL/tempo, ~ diag(2):id(12)),
              rcov=~ Especies:Idades:Lotes:Grupos:corgh(Tempos),
              aom=T,
              data= Euc.Pulp2x2.dat, start.values="gamm.txt")
gamm.constr <- gamm$gammas.table[c(21:26),]
gamm.constr$fac <- factor(rep(1:6, times=1))
G <- asreml.constraints(~ fac, gammas=gamm.constr)
mod10.corgh<- asreml(peso ~ Especies*Idades*tempo,
                    random=~ spl(tempo) +
                          dev(tempo):Idades +
                          str(~ EIL/tempo, ~ diag(2):id(12)),
                    rcov=~ Especies:Idades:Lotes:Grupos:corgh(Tempos),
                    aom=T,
                    data= Euc.Pulp2x2.dat, constraint=G, G.param="gamm.txt")
summary(mod10.corgh)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.idh)
info.crit(mod10.corgh) # fico com mod10.idh
graficos(mod10.corgh, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.corgh, mod10.idh, positive.zero=F)

tapply(Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$Especies, var)
tapply(Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$Idades, var)
tapply(Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$Lotes, var)
tapply(Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EI, var)

# variancias residuais diferentes para L
mod10.idh.atL<- asreml(peso ~ Especies*Idades*tempo,
                      random=~ spl(tempo) +
                            dev(tempo):Idades +
                            str(~ EIL/tempo, ~ diag(2):id(12)) +
                            str(~ EILG/tempo, ~ diag(2):id(48)),
                      rcov=~ Especies:Idades:at(Lotes):Grupos:idh(Tempos),
                      aom=T,
                      data= Euc.Pulp2x2.dat)
summary(mod10.idh.atL)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.idh)
info.crit(mod10.idh.atL)
graficos(mod10.idh.atL, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

```

```

reml.lrt(mod10.idh.atL, mod10.idh, positive.zero=F) # fico com mod10.idh

# variancias residuais diferentes para a combinao de EI
mod10.idh.atEI<- asreml(peso ~ Especies*Idades*tempo,
                        random=~ spl(tempo) +
                              dev(tempo):Idades +
                              str(~ EIL/tempo, ~ diag(2):id(12)) +
                              str(~ EILG/tempo, ~ diag(2):id(48)),
                        rcov=~ at(EI):Lotes:Grupos:idh(Tempos),
                        aom=T,
                        data= Euc.Pulp2x2.dat)
summary(mod10.idh.atEI)$varcomp
info.crit(mod2.1)
info.crit(mod10)
info.crit(mod10.idh)
info.crit(mod10.idh.atEI)
graficos(mod10.idh.atEI, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh.atEI, mod10.idh, positive.zero=F) # fico com mod10.idh

# EILG int-coef + EILG componente
mod10.idh3<- asreml(peso ~ Especies*Idades*tempo,
                    random=~ spl(tempo) +
                          dev(tempo):Idades +
                          str(~ EIL/tempo, ~ diag(2):id(12)) +
                          EILG,
                    rcov=~ Especies:Idades:Lotes:Grupos:idh(Tempos),
                    aom=T,
                    data= Euc.Pulp2x2.dat)
summary(mod10.idh3)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh3)
graficos(mod10.idh3, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh, mod10.idh3, positive.zero=T) # fico com mod10.idh

# - spl(tempo)
mod10.idh5<- asreml(peso ~ Especies*Idades*tempo,
                    random=~ dev(tempo):Idades +
                          str(~ EIL/tempo, ~ diag(2):id(12)) +
                          str(~ EILG/tempo, ~ diag(2):id(48)),
                    rcov=~ Especies:Idades:Lotes:Grupos:idh(Tempos),
                    aom=T,
                    data= Euc.Pulp2x2.dat)
summary(mod10.idh)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
graficos(mod10.idh5, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh, mod10.idh5, positive.zero=F) # fico com mod10.idh5

```

```

# teste wald-F
wald.asreml(mod10.idh5, denDF="algebraic", ssType="conditional")

mod10.idh5.fix<- asreml(peso ~ Especies + tempo/Idades,
                      random=~ dev(tempo):Idades +
                          str(~ EIL/tempo, ~ diag(2):id(12)) +
                          str(~ EILG/tempo, ~ diag(2):id(48)),
                      rcov=~ Especies:Idades:Lotes:Grupos:indh(Tempos),
                      aom=T,
                      data= Euc.Pulp2x2.dat)
wald.asreml(mod10.idh5.fix, denDF="algebraic", ssType="conditional")
summary(mod10.idh5.fix)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
info.crit(mod10.idh5.fix)
graficos(mod10.idh5.fix, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

# + spl(tempo):Idades
mod10.idh5.fixB0<- asreml(peso ~ Especies + tempo/Idades,
                        random=~ spl(tempo):Idades +
                            dev(tempo):Idades +
                            str(~ EIL/tempo, ~ diag(2):id(12)) +
                            str(~ EILG/tempo, ~ diag(2):id(48)),
                        rcov=~ Especies:Idades:Lotes:Grupos:indh(Tempos),
                        aom=T,
                        data= Euc.Pulp2x2.dat)
summary(mod10.idh5.fixB0)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
info.crit(mod10.idh5.fix)
info.crit(mod10.idh5.fixB0)
graficos(mod10.idh5.fixB0, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh5.fixB0, mod10.idh5.fix, positive.zero=T) # fico com mod10.idh5.fixB0

# - dev(tempo):Idades
mod10.idh5.fixB1<- asreml(peso ~ Especies + tempo/Idades,
                        random=~ spl(tempo):Idades +
                            str(~ EIL/tempo, ~ diag(2):id(12)) +
                            str(~ EILG/tempo, ~ diag(2):id(48)),
                        rcov=~ Especies:Idades:Lotes:Grupos:indh(Tempos),
                        aom=T,
                        data= Euc.Pulp2x2.dat)
summary(mod10.idh5.fixB1)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
info.crit(mod10.idh5.fixB0)
info.crit(mod10.idh5.fixB1)
graficos(mod10.idh5.fixB1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),

```

```

      Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh5.fixB0, mod10.idh5.fixB1, positive.zero=T) # fico com mod10.idh5.fixB1

# - EILG int-coef + EILG componente
mod10.idh5.fixB2<- asreml(peso ~ Especies + tempo/Idades,
      random=~ spl(tempo):Idades +
      str(~ EIL/tempo, ~ diag(2):id(12))+
      EILG,
      rcov=~ Especies:Idades:Lotes:Grupos:idh(Tempos),
      aom=T,
      data= Euc.Pulp2x2.dat)
summary(mod10.idh5.fixB2)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
info.crit(mod10.idh5.fixB0)
info.crit(mod10.idh5.fixB1)
info.crit(mod10.idh5.fixB2)
graficos(mod10.idh5.fixB2, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
      Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh5.fixB1, mod10.idh5.fixB2, positive.zero=T) # fico com mod10.idh5.fixB1

# - EIL int-coef - EILG int-coef + EIL componente
mod10.idh5.fixB3<- asreml(peso ~ Especies + tempo/Idades,
      random=~ spl(tempo):Idades +
      EILG,
      rcov=~ Especies:Idades:Lotes:Grupos:idh(Tempos),
      aom=T,
      data= Euc.Pulp2x2.dat)
summary(mod10.idh5.fixB3)$varcomp
info.crit(mod2.1)
info.crit(mod10.idh)
info.crit(mod10.idh5)
info.crit(mod10.idh5.fixB0)
info.crit(mod10.idh5.fixB1)
info.crit(mod10.idh5.fixB3)
graficos(mod10.idh5.fixB3, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
      Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)
reml.lrt(mod10.idh5.fixB1, mod10.idh5.fixB3, positive.zero=T) # fico com mod10.idh5.fixB1

#teste wald-F
wald.asreml(mod10.idh5.fixB1, denDF="algebraic", ssType="conditional")

mod10.idh5.fixB<- asreml(peso ~ (Especies + Idades)*tempo,
      random=~ spl(tempo):Idades +
      EILG,
      rcov=~ Especies:Idades:Lotes:Grupos:idh(Tempos),
      aom=T,
      data= Euc.Pulp2x2.dat)
wald.asreml(mod10.idh5.fixB, denDF="algebraic", ssType="conditional")
summary(mod10.idh5.fixB)$varcomp
info.crit(mod2.1)

```

```

info.crit(mod10.idh5.fixB1)
info.crit(mod10.idh5.fixB)
graficos(mod10.idh5.fixB, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$EILG)

pv <- predict(mod10.idh5.fixB, classify = "Especies:Idades:tempo",
              predictpoints=list(tempo=seq(30,180,5)))
pred <- pv$predictions$pvals
pred$Idades <- factor(pred$Idades)
Especies.Idades <- as.factor(fac.combine(list(pred$Especies, pred$Idades)))
levels(pred$Especies)
levels(Especies.Idades) <- c("5:E. alba", "7:E. alba",
                             "5:E. saligna", "7:E. saligna")
names(pred)

X11()
par(mfrow=c(2,2))
plot(peso ~ tempo, data=Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Especies=="E. alba" &
                                         Euc.Pulp2x2.dat$Idades==5,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005))
lines(predicted.value ~ tempo, data=pred[pred$Especies=="E. alba" &
                                          pred$Idades==5,], col=2, lwd=2)
#
plot(peso ~ tempo, data=Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Especies=="E. saligna" &
                                         Euc.Pulp2x2.dat$Idades==5,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005))
lines(predicted.value ~ tempo, data=pred[pred$Especies=="E. saligna" &
                                          pred$Idades==5,], col=2, lwd=2)
#
plot(peso ~ tempo, data=Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Especies=="E. alba" &
                                         Euc.Pulp2x2.dat$Idades==7,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005))
lines(predicted.value ~ tempo, data=pred[pred$Especies=="E. alba" &
                                          pred$Idades==7,], col=2, lwd=2)
#
plot(peso ~ tempo, data=Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Especies=="E. saligna" &
                                         Euc.Pulp2x2.dat$Idades==7,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005))
lines(predicted.value ~ tempo, data=pred[pred$Especies=="E. saligna" &
                                          pred$Idades==7,], col=2, lwd=2)

X11()
plot(predicted.value ~ tempo, data=pred[pred$Especies=="E. alba" &
                                         pred$Idades==5,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005),
      lty=1, type="l", ylab="Peso específico aparente (g/cm3)",
      xlab="tempo de refinação (min)")
lines(predicted.value ~ tempo, data=pred[pred$Especies=="E. alba" &
                                         pred$Idades==7,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005),
      lty=2)

```

```
lines(predicted.value ~ tempo, data=pred[pred$Especie=="E. saligna" &
                                           pred$Idades==5,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005),
      lty=3)
lines(predicted.value ~ tempo, data=pred[pred$Especie=="E. saligna" &
                                           pred$Idades==7,],
      ylim=c(min(Euc.Pulp2x2.dat$peso)-.005, max(Euc.Pulp2x2.dat$peso)+.005),
      lty=4)
legend(125, 0.85, c("E. alba:5", "E. alba:7",
                    "E. saligna:5", "E. saligna:7"), lty=c(1:4))
```


ANEXO D - Códigos para ajuste do modelo para os dados do experimento com polpas de Eucalipto reunidas as Espécies E. alba com 5 anos e E. saligna com 7 anos

```

Euc.Pulp2x2.dat <- asreml.read.table("peso_novo.csv", header=T, dec="," ,sep=";")
Euc.Pulp2x2.dat$TL <- as.factor(fac.combine(list(Euc.Pulp2x2.dat$Trat,
                                                Euc.Pulp2x2.dat$L)))
Euc.Pulp2x2.dat$TLG <- as.factor(fac.combine(list(Euc.Pulp2x2.dat$Trat,
                                                Euc.Pulp2x2.dat$L,
                                                Euc.Pulp2x2.dat$Grupos)))

X11()
par(mfrow=c(1,3))
###
T1<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==1],
            Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Trat==1], mean)
plot(T1, ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                max(na.omit(Euc.Pulp2x2.dat$peso))+.005),
      xlab="tempo de refinação (min)",ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. alba:5 e E. saligna:7", type="l", axes=F,
      col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                max(na.omit(Euc.Pulp2x2.dat$peso))+.005, length=8),
      labels =seq(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                  max(na.omit(Euc.Pulp2x2.dat$peso))+.005, , length=8))
box()
for(i in 1:24){
  lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==1][(6*i-5):(6*i)],, lty=2)
  points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==1][(6*i-5):(6*i)])
}
###
T2<- tapply(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==2][1:72],
            Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Trat==2][1:72], mean)
plot(T2, ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                max(na.omit(Euc.Pulp2x2.dat$peso))+.005),
      xlab="tempo de refinação (min)",ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. alba:7", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                max(na.omit(Euc.Pulp2x2.dat$peso))+.005, length=8),
      labels =seq(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                  max(na.omit(Euc.Pulp2x2.dat$peso))+.005, , length=8))
box()
for(i in 1:24){
  lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==2][(6*i-5):(6*i)],, lty=2)
  points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==2][(6*i-5):(6*i)])
}
###
T3<- tapply(na.omit(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==3]),
            Euc.Pulp2x2.dat$tempo[Euc.Pulp2x2.dat$Trat==3][1:72], mean)
plot(T3, ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))- .005,
                max(na.omit(Euc.Pulp2x2.dat$peso))+.005),

```

```

      xlab="tempo de refinação (min)",ylab="Peso específico aparente (g/cm3)",
      main="Espécies:Idades : E. saligna:5", type="l", axes=F, col="red", lwd=2)
axis(1, at=seq(1,6,1), labels =seq(30,180,length=6))
axis(2, at=seq(min(na.omit(Euc.Pulp2x2.dat$peso))-0.005,
      max(na.omit(Euc.Pulp2x2.dat$peso))+0.005, length=8),
      labels =seq(min(na.omit(Euc.Pulp2x2.dat$peso))-0.005,
      max(na.omit(Euc.Pulp2x2.dat$peso))+0.005, , length=8))
box()
for(i in 1:24){
lines(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==3][(6*i-5):(6*i)],, lty=2)
points(Euc.Pulp2x2.dat$peso[Euc.Pulp2x2.dat$Trat==3][(6*i-5):(6*i)])
}

mod1 <- asreml(peso ~ Trat*Tempos,
              random= ~ Trat:L/Tempos + Trat:L:Grupos,
              rcov=~ Trat:L:Grupos:Tempos,
              aom=T,
              data=Euc.Pulp2x2.dat)

summary(mod1)$varcomp
info.crit(mod1)
#graficos(mod1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
#          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)

mod2<- asreml(peso ~ Trat*tempo,
              random=~ spl(tempo)/Trat +
              dev(tempo)/Trat +
              str(~ TL/tempo, ~ diag(2):id(18)) +
              str(~ TLG/tempo, ~ diag(2):id(72)) +
              TL:(Tempos + spl(tempo)) +
              TLG:(spl(tempo)),
              rcov=~ Trat:L:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod2)$varcomp
info.crit(mod1)
info.crit(mod2)
graficos(mod2, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)

# removendo os termos de fronteira
mod2.1<- asreml(peso ~ Trat*tempo,
              random=~ spl(tempo)/Trat +
              str(~ TL/tempo, ~ diag(2):id(18)) +
              str(~ TLG/tempo, ~ diag(2):id(72)) +
              TLG:(spl(tempo)),
              rcov=~ Trat:L:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod2.1)$varcomp
info.crit(mod1)
info.crit(mod2.1)
graficos(mod2.1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)

```



```

rcov=~ Trat:L:Grupos:Tempos,
aom=T,
data= Euc.Pulp2x2.dat)

summary(mod6)$varcomp
info.crit(mod1)
info.crit(mod4)
info.crit(mod6)
graficos(mod6, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod6, mod4, positive.zero=F) # fico com mod6

# - TL lin(tempo) + TL componente
mod7<- asreml(peso ~ Trat*tempo,
              random=~ spl(tempo) +
                    TL +
                    str(~ TLG/tempo, ~ us(2):id(72)),
              rcov=~ Trat:L:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod7)$varcomp
info.crit(mod1)
info.crit(mod6)
info.crit(mod7)
graficos(mod7, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod6, mod7, positive.zero=T) # fico com mod7

# - spl(tempo)
mod8<- asreml(peso ~ Trat*tempo,
              random=~ TL +
                    str(~ TLG/tempo, ~ us(2):id(72)),
              rcov=~ Trat:L:Grupos:Tempos,
              aom=T,
              data= Euc.Pulp2x2.dat)

summary(mod8)$varcomp
info.crit(mod1)
info.crit(mod7)
info.crit(mod8)
graficos(mod8, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7, mod8, positive.zero=T) # fico com mod7

# modificando R

mod7.ar1<- asreml(peso ~ Trat*tempo,
                  random=~ spl(tempo) +
                        TL +
                        str(~ TLG/tempo, ~ us(2):id(72)),
                  rcov=~ Trat:L:Grupos:ar1(Tempos),
                  aom=T,
                  data= Euc.Pulp2x2.dat)

summary(mod7.ar1)$varcomp
info.crit(mod1)

```

```

info.crit(mod7)
info.crit(mod7.ar1)
graficos(mod7.ar1, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7.ar1, mod7, positive.zero=F) # fico com mod7

mod7.idh<- asreml(peso ~ Trat*tempo,
                  random=~ spl(tempo) +
                        TL +
                        str(~ TLG/tempo, ~ diag(2):id(72)),
                  rcov=~ Trat:L:Grupos:idh(Tempos),
                  aom=T,
                  data= Euc.Pulp2x2.dat)
summary(mod7.idh)$varcomp
info.crit(mod1)
info.crit(mod7.idh)
graficos(mod7.idh, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7.idh, mod7, positive.zero=F) # fico com mod7.idh

mod7.ar1h<- asreml(peso ~ Trat*tempo,
                   random=~ spl(tempo) +
                         TL +
                         str(~ TLG/tempo, ~ diag(2):id(72)),
                   rcov=~ Trat:L:Grupos:ar1h(Tempos),
                   aom=T,
                   data= Euc.Pulp2x2.dat)
summary(mod7.ar1h)$varcomp
info.crit(mod1)
info.crit(mod7.ar1h)
graficos(mod7.ar1h, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7.ar1h, mod7.idh, positive.zero=F) # fico com mod7.idhh

mod7.corgh<- asreml(peso ~ Trat*tempo,
                    random=~ spl(tempo) +
                          TL,
                    rcov=~ Trat:L:Grupos:corgh(Tempos),
                    aom=T,
                    data= Euc.Pulp2x2.dat)
summary(mod7.corgh)$varcomp
info.crit(mod1)
info.crit(mod7.idh)
info.crit(mod7.ar1h)
info.crit(mod7.corgh)
graficos(mod7.corgh, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7.corgh, mod7.idh, positive.zero=F) # fico com mod7.corgh

# + spl(tempo):Trat
mod7.corgh2<- asreml(peso ~ Trat*tempo,
                     random=~ spl(tempo)/TL +
                           TL,

```

```

rcov=~ Trat:L:Grupos:corgh(Tempos),
aom=T,
data= Euc.Pulp2x2.dat)
summary(mod7.corgh2)$varcomp
info.crit(mod1)
info.crit(mod7.idh)
info.crit(mod7.corgh)
info.crit(mod7.corgh2)
graficos(mod7.corgh2, Euc.Pulp2x2.dat, long=c("tempo"), Long=c("Tempos"),
          Euc.Pulp2x2.dat$peso, Euc.Pulp2x2.dat$TLG)
reml.lrt(mod7.corgh2, mod7.corgh, positive.zero=T) # fico com mod7.corgh

#teste Wald-F
wald.asreml(mod7.corgh, denDF="algebraic", ssType="conditional")

pv <- predict(mod7.corgh, classify = "Trat:tempo",
              predictpoints=list(tempo=seq(30,180,5)))
pred <- pv$predictions$pvals
pred$Trat <- factor(pred$Trat)
names(pred)

X11()
par(mfrow=c(1,3))
plot(peso ~ tempo, data=na.omit(Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Trat==1,]),
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005),
     main="E. alba:5 e E. saligna:7")
lines(predicted.value ~ tempo, data=pred[pred$Trat==1,], col=2, lwd=2)
#
plot(peso ~ tempo, data=na.omit(Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Trat==2,]),
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005), main="E. alba:7")
lines(predicted.value ~ tempo, data=pred[pred$Trat==2,], col=2, lwd=2)
#
plot(peso ~ tempo, data=na.omit(Euc.Pulp2x2.dat[Euc.Pulp2x2.dat$Trat==3,]),
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005), main="E. saligna:5")
lines(predicted.value ~ tempo, data=pred[pred$Trat==3,], col=2, lwd=2)

X11()
plot(predicted.value ~ tempo, data=pred[pred$Trat==1,],
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005), lty=1, type="l",
     ylab="Peso específico aparente (g/cm3)", xlab="tempo de refinação (min)")
lines(predicted.value ~ tempo, data=pred[pred$Trat==2,],
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005), lty=2)
lines(predicted.value ~ tempo, data=pred[pred$Trat==3,],
     ylim=c(min(na.omit(Euc.Pulp2x2.dat$peso))-.005,
            max(na.omit(Euc.Pulp2x2.dat$peso))+.005), lty=3)

legend(115, 0.85, c("E. alba:5 e E. saligna:7",
                    "E. alba:7", "E. saligna:5"), lty=c(1:3))

```