Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

Modelos lineares mistos para dados longitudinais em ensaio fatorial com tratamento adicional

Gilson Silvério da Rocha

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

Piracicaba 2015 Gilson Silvério da Rocha Licenciado em Matemática

Modelos lineares mistos para dados longitudinais em ensaio fatorial com tratamento adicional

versão revisada de acordo com a resolução CoPG
r6018 de 2011

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 Experimentação Agronômica

Piracicaba 2015

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

Rocha, Gilson Silvério da

Modelos lineares mistos para dados longitudinais em ensaio fatorial com tratamento adicional / Gilson Silvério da Rocha. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015. 110 p. : il.

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

1. Efeitos aleatórios 2. Teste da razão de verossimilhanças 3. Diagrama de Hasse 4. Regressão 5. Matéria seca 6. Ulexita 7. Ácido bórico 8. Alfafa I. Título

> CDD 519.5 R672m

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

DEDICATÓRIA

A minha mãe

Sueli,

por ser a melhor mãe do mundo.

A minha noiva

Simone,

"I feel blessed to be loved by someone as wonderful as you."

Com amor, DEDICO.

AGRADECIMENTOS

A Deus, pelo dom da vida e por sempre iluminar meu caminho, dando-me forças para vencer todos os obstáculos e para tornar possível mais este sonho.

À Escola Superior de Agricultura "Luiz de Queiroz" da Universidade de São Paulo e ao Programa de Pós-Graduação em Estatística e Experimentação Agronômica, pela oportunidade de realizar o curso.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agronômica, em especial àqueles que contribuíram para minha formação acadêmica.

A minha mãe Sueli, pelo amor incondicional, pela educação, pela confiança, pelo cuidado e por estar sempre presentes em todas as etapas da minha vida.

Ao meu irmão, Gilmar, pela amizade, pela confiança e por estar sempre disposto a ajudar.

A minha noiva Simone por me apoiar em todos os momentos.

À professora e orientadora Dr^a Sônia Maria De Stefano Piedade por ter me "adotado" e por todos os ensinamentos.

À professora Dr^a Renata Alcarde Sermarini, pelo exemplo de profissional, pela amizade, pelos sábios ensinamentos, responsáveis pelo meu crescimento profissional, e por estar sempre disponível e atenciosa para me atender, mesmo com várias outras obrigações e compromissos.

Às pesquisadoras Dr^a Maria Cristina Neves de Oliveira e Dr^a Ivani de Oliveira Negrão Lopes, pelo apoio, por sempre estarem dispostas a ajudar e por permitirem a realização de mais esse sonho.

Ao pesquisador Dr Adônis Moreira, pelo apoio e pela concessão dos dados utilizados na pesquisa.

Ao Dr Leonardo Cesar Ferreira, por toda a ajuda, pela amizade, pelos ensinamentos e por sempre estar presente.

Ao professor Dr Fabyano Fonseca e Silva pelo exemplo de profissional, pela amizade e por sempre ajudar, mesmo com várias outras obrigações e compromissos.

Ao professor Dr César Gonçalves de Lima, pela ajuda e pelas sugestões para o enriquecimento deste trabalho.

A todos os amigos do doutorado, pela companhia nos estudos e no lazer. O

apoio de vocês foi essencial.

- A todos os amigos da Embrapa Soja, pelo apoio nas horas mais difíceis.
- À CAPES, pela concessão da bolsa de estudos.

A todos que, de alguma forma, contribuíram para o meu crescimento profissional e para a realização deste trabalho.

"Jamais considere seus estudos como uma obrigação, mas como uma oportunidade invejável para aprender a conhecer a influência libertadora da beleza do reino do espírito, para seu próprio prazer pessoal e para proveito da comunidade à qual seu futuro trabalho pertencer."

7

Albert Einstein

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Medidas repetidas	21
2.2 Modelos lineares mistos	24
2.2.1 Estruturas das matrizes de covariâncias	29
2.2.2 Especificações de estruturas de covariâncias	33
2.2.3 Modelo linear misto marginal	39
2.2.4 Estimação em modelos lineares mistos	40
2.2.4.1 Máxima Verossimilhança - MV	41
2.2.4.2 Máxima Verossimilhança Restrita - MVR	44
2.2.5 Predição em modelos lineares mistos	45
2.2.6 Estimação e predição pelas equações de modelos mistos	46
2.2.7 Diagrama de Hasse na construção de modelos lineares mistos	47
2.2.8 Seleção de modelos	48
2.2.8.1 Teste da Razão de Verossimilhanças	49
2.2.8.2 Testes adicionais para parâmetros de efeitos fixos	51
2.2.8.3 Critérios de informação	53
2.2.8.4 Estratégias para construção de modelos	55
2.2.9 Diagnósticos	57
2.2.9.1 Resíduos	58
2.3 A cultura da alfafa	60
3 MATERIAL E MÉTODOS	63
3.1 Material	63
3.2 Métodos	64
3.2.1 Identificação dos tratamentos	64
3.2.2 Análise exploratória	65
$3.2.3$ Diagrama de Hasse e estratégia $top{-}down$ na especificação do modelo estatístico	65

3.2.4 Seleção dos efeitos aleatórios	66
3.2.5 Seleção da estrutura de covariâncias para os resíduos	68
3.2.6 Teste para efeitos fixos	69
3.2.7 Estudo de regressão	69
3.2.7.1 Diagnósticos	70
4 RESULTADOS E DISCUSSÃO	71
4.1 Análise exploratória	71
4.2 Diagrama de Hasse e número de graus de liberdade	75
4.3 Seleção dos efeitos aleatórios	78
4.4 Seleção da estrutura de covariâncias para os resíduos	79
4.5 Teste para efeitos fixos	80
4.6 Estudo de regressão	81
4.6.1 Diagnósticos	84
4.7 Comparação de tratamentos	85
5 CONCLUSÃO	87
REFERÊNCIAS	89
ANEXOS	95

RESUMO

Modelos lineares mistos para dados longitudinais em ensaio fatorial com tratamento adicional

Em experimentos agronômicos são comuns ensaios planejados para estudar determinadas culturas por meio de múltiplas mensurações realizadas na mesma unidade amostral ao longo do tempo, espaço, profundidade entre outros. Essa forma com que as mensurações são coletadas geram conjuntos de dados que são chamados de dados longitudinais. Nesse contexto, é de extrema importância a utilização de metodologias estatísticas que sejam capazes de identificar possíveis padrões de variação e correlação entre as mensurações. A possibilidade de inclusão de efeitos aleatórios e de modelagem das estruturas de covariâncias tornou a metodologia de modelos lineares mistos uma das ferramentas mais apropriadas para a realização desse tipo de análise. Entretanto, apesar de todo o desenvolvimento teórico e computacional, a utilização dessa metodologia em delineamentos mais complexos envolvendo dados longitudinais e tratamentos adicionais, como os utilizados na área de forragicultura, ainda é passível de estudos. Este trabalho envolveu o uso do diagrama de Hasse e da estratégia top-down na construção de modelos lineares mistos no estudo de cortes sucessivos de forragem provenientes de um experimento de adubação com boro em alfafa (Medicago sativa L.) realizado no campo experimental da Embrapa Pecuária Sudeste. Primeiramente, considerou-se uma abordagem qualitativa para todos os fatores de estudo e devido à complexidade do delineamento experimental optou-se pela construção do diagrama de Hasse. A incorporação de efeitos aleatórios e seleção de estruturas de covariâncias para os resíduos foram realizadas com base no teste da razão de verossimilhanças calculado a partir de parâmetros estimados pelo método da máxima verossimilhança restrita e nos critérios de informação de Akaike (AIC), Akaike corrigido (AICc) e bayesiano (BIC). Os efeitos fixos foram testados por meio do teste Wald-F e, devido aos efeitos significativos das fontes de variação associadas ao fator longitudinal, desenvolveu-se um estudo de regressão. A construção do diagrama de Hasse foi fundamental para a compreensão e visualização simbólica do relacionamento de todos os fatores presentes no estudo, permitindo a decomposição das fontes de variação e de seus graus de liberdade, garantindo que todos os testes fossem realizados corretamente. A inclusão de efeito aleatório associado à unidade experimental foi essencial para a modelagem do comportamento de cada unidade e a estrutura de componentes de variância com heterogeneidade, incorporada aos resíduos, foi capaz de modelar eficientemente a heterogeneidade de variâncias presente nos diferentes cortes da cultura da alfafa. A verificação do ajuste foi realizada por meio de gráficos de diagnósticos de resíduos. O estudo de regressão permitiu avaliar a produtividade de matéria seca da parte aérea da planta (kg ha $^{-1}$) de cortes consecutivos da cultura da alfafa, envolvendo a comparação de adubações com diferentes fontes e doses de boro. Os melhores resultados de produtividade foram observados para a combinação da fonte ulexita com as doses 3, 6 e 9 kg ha⁻¹ de boro.

Palavras-chave: Efeitos aleatórios; Teste da razão de verossimilhanças; Diagrama de Hasse; Regressão; Matéria seca; Ulexita; Ácido bórico; Alfafa

ABSTRACT

Mixed linear models for longitudinal data in a factorial experiment with additional treatment

Assays aimed at studying some crops through multiple measurements performed in the same sample unit along time, space, depth etc. have been frequently adopted in agronomical experiments. This type of measurement originates a dataset named longitudinal data, in which the use of statistical procedures capable of identifying possible standards of variation and correlation among measurements has great importance. The possibility of including random effects and modeling of covariance structures makes the methodology of mixed linear models one of the most appropriate tools to perform this type of analysis. However, despite of all theoretical and computational development, the use of such methodology in more complex designs involving longitudinal data and additional treatments, such as those used in forage crops, still needs to be studied. The present work covered the use of the Hasse diagram and the top-down strategy in the building of mixed linear models for the study of successive cuts from an experiment involving boron fertilization in alfalfa (Medicago sativa L.) carried out in the field area of Embrapa Southeast Livestock. First, we considered a qualitative approach for all study factors and we chose the Hasse diagram building due to the model complexity. The inclusion of random effects and selection of covariance structures for residues were performed based on the likelihood ratio test, calculated based on parameters estimated through the restricted maximum likelihood method, the Akaike's Information Criterion (AIC), the Akaike's information criterion corrected (AICc) and the Bayesian Information Criterion (BIC). The fixed effects were analyzed through the Wald-F test and we performed a regression study due to the significant effects of the variation sources associated with the longitudinal factor. The Hasse diagram building was essential for understanding and symbolic displaying regarding the relation among all factors present in the study, thus allowing variation sources and their degrees of freedom to be decomposed, assuring that all tests were correctly performed. The inclusion of random effect associated with the sample unit was essential for modeling the behavior of each unity. Furthermore, the structure of variance components with heterogeneity, added to the residues, was capable of modeling efficiently the heterogeneity of variances present in the different cuts of alfalfa plants. The fit was checked by residual diagnostic plots. The regression study allowed us to evaluate the productivity of shoot dry matter (kg ha⁻¹) related to successive cuts of alfalfa plants, involving the comparison of fertilization with different boron sources and doses. We observed the best productivity in the combination of the source ulexite with the doses 3, 6 and 9 kg ha^{-1} boron.

Keywords: Random effects; Likelihood ratio test; Hasse diagram; Regression; Dry matter; Ulexite; Boric acid; Alfalfa

LISTA DE FIGURAS

Figura 1 -	Ensaio com a cultura da alfafa realizado no campo experimental da	
	Embrapa Pecuária Sudeste	63
Figura 2 -	Perfis individuais da produção de matéria seca da parte a érea (kg $\mathrm{ha}^{-1})$	
	de plantas de alfafa, por unidade experimental ao longo dos dias de	
	avaliação	71
Figura 3 -	Perfis médios da produção de matéria seca da parte a érea (kg $\mathrm{ha}^{-1})$ de	
	plantas de alfafa, por tratamento ao longo dos dias de avaliação	72
Figura 4 -	Perfis médios \pm erro padrão da produção de matéria seca da parte a érea	
	(kg ha^{-1}) de plantas de alfafa por doses de boro $(0, 1, 3, 6 \text{ e 9 kg ha}^{-1})$	
	ao longo dos dias de avaliação. Fontes: (a) controle, (b) ácido bórico	
	$(H_3BO_3), (c)$ ulexita	73
Figura 5 -	Perfis médios \pm erro padrão da produção de matéria seca da parte a érea	
	$({\rm kg\ ha^{-1}})$ de plantas de alfafa por fontes de boro (controle, ácido bórico	
	e ulexita) ao longo dos dias de avaliação. Doses: (a) 0 kg ha $^{-1}$, (b) 1 kg	
	ha ⁻¹ , (c) 3 kg ha ⁻¹ , (d) 6 kg ha ⁻¹ , 9 kg ha ⁻¹	74
Figura 6 -	Diagramas de Hasse com indicação para obtenção dos graus de liberdade	
	(a) e com os valores numéricos dos graus de liberdade (b), envolvendo	
	fatores não aleatorizados do experimento de adubação com boro em alfafa	75
Figura 7 -	Diagrama de Hasse com indicação para obtenção dos graus de liberdade	
	envolvendo fatores aleatorizados do experimento de adubação com boro	
	em alfafa	76
Figura 8 -	Diagrama de Hasse com os valores numéricos dos graus de liberdade	
	envolvendo fatores aleatorizados do experimento de adubação com boro	
	em alfafa	77
Figura 9 -	Diagnóstico para o modelo (37)+UN(1): (a) Resíduos condicionais es-	
	tudentizados em relação ao valores ajustados; (b) Resíduos condicionais	
	estudentizados em função dos dias de avaliação; (c) Gráfico de quantil-	
	quantil para os resíduos condicionais estudentizados; (d) Valores ajusta-	
	dos vs observados	84

LISTA DE TABELAS

Tabela 1 -	Composição dos tratamentos no experimento de adubação com boro em	
	alfafa	64
Tabela 2 -	Decomposição dos números de graus de liberdade para o experimento	
	de adubação com boro em alfafa	78
Tabela 3 -	Resultados dos testes de hipóteses envolvendo modelos com diferentes	
	partes aleatórias ajustados na análise da matéria seca para o experimento	
	de adubação com boro	79
Tabela 4 -	Síntese das estatísticas obtidas nos testes envolvendo o modelo linear	
	misto (34) com diferentes estruturas de correlação para a matriz de erros	
	intra-unidades experimentais ajustados na análise da matéria seca para	
	o experimento de adubação com boro	80
Tabela 5 -	Teste Wald- F para os efeitos fixos do modelo linear misto (34) com estru-	
	tura de componentes de variância com heterogeneidade (" $Banded\ Main$	
	Diagonal - $UN(1)$ ") para a matriz de erros intra-unidades experimentais	
	(\boldsymbol{R}_i)	81
Tabela 6 -	Teste Wald- F para os efeitos fixos do modelo (37) com estrutura de	
	componentes de variância ("Variance Components - VC") para a matriz	
	de erros intra-unidades experimentais (\boldsymbol{R}_i)	82
Tabela 7 -	Síntese das estatísticas obtidas nos testes envolvendo o modelo linear	
	misto (37) com diferentes estruturas de correlação para a matriz de erros	
	intra-unidades experimentais ajustados na análise da matéria seca para	
	o experimento de adubação com boro	82
Tabela 8 -	Teste Wald- F para os efeitos fixos do modelo (37) com estrutura de	
	componentes de variância com heterogeneidade (" $Banded\ Main\ Diagonal$	
	- $\mathit{UN(1)^{"}}$ para a matriz de erros intra-unidades experimentais $(oldsymbol{R}_i)$	83

1 INTRODUÇÃO

As diferentes áreas de pesquisas científicas, como as ciências agrárias, ambientais, médicas, biológicas e sociais, necessitam de suporte metodológico estatístico para o desenvolvimento e viabilização de atividades relacionadas ao planejamento de pesquisas, análise de dados, interpretação e apresentação dos resultados. Nestas áreas, existem pesquisas envolvendo conjuntos de dados agrupados por um ou mais fatores. Na agricultura, por exemplo, os dados podem estar agrupados por áreas de terra, épocas de plantio e tipos de adubação. Dados coletados de espécies florestais em matas, de pacientes em clínicas, de estudantes em salas de aulas são exemplos para as outras áreas.

Existem também conjuntos de dados gerados por meio de medidas repetidas realizadas em uma mesma unidade experimental. Essas medidas podem ser coletadas sob diferentes condições experimentais ou observacionais. Particularmente, quando essas medidas são coletadas ordenadamente ao longo de uma dimensão específica (tempo ou espaço, por exemplo) o conjunto de medidas repetidas recebe o nome de dados longitudinais (WEST; WELCH; GALECKI, 2007).

Na área de forragicultura, por exemplo, são realizados vários cortes sucessivos de forragens na mesma unidade experimental ao decorrer do ciclo da cultura e com o material coletado em cada corte são realizadas mensurações para caraterísticas de interesse, tais como teor de boro, produção de matéria seca, teor de nitrato, teor de proteína, fibra em detergente neutro etc. É de grande interesse estudar o conjunto de dados obtidos a partir das múltiplas mensurações realizadas ao longo do tempo, uma vez que esse estudo permite avaliar a eficiência dos tratamentos testados ao longo de todo o ciclo, possibilitando a compreensão da dinâmica de todo o processo de avaliação, não se restringindo a resultados de análises realizadas separadamente com os dados de cada corte. Contudo, é fundamental a utilização de uma metodologia estatística adequada para esse tipo de estudo. Segundo Littell, Henry e Ammerman (1998), é preciso investigar se existe correlação entre as mensurações da mesma unidade em tempos distintos e se a variabilidade em cada tempo pode ser considerada homogênea ou se necessita ser modelada.

No decorrer dos anos, surgiram várias abordagens para se analisar dados longitudinais. Porém, a metodologia envolvendo a especificação de modelos lineares mistos proposta por Laid e Ware (1982), utilizando as ideias de Harville (1977), destacouse entre as demais e vem sendo disseminada por diversos autores (LITTELL; HENRY; AMMERMAN, 1998; LITTELL; PENDERGAST; NATARAJAN, 2000; LITTELL et al., 2006; PINHEIRO; BATES, 2000; VERBEKE; MOLENBERGHS, 2000; WEST; WELCH; GALECKI, 2007).

Além de incluir efeitos fixos como no modelo linear padrão, o modelo linear misto incorpora efeitos aleatórios capazes de modelar adequadamente a correlação e a variabilidade dentro e entre unidades experimentais. Com um modelo muito mais abrangente, a metodologia por modelos mistos permite analisar adequadamente o padrão de relacionamento gerado pela estrutura de dados longitudinais. Contudo, embora a utilização dessa metodologia seja uma poderosa ferramenta estatística, a formulação do modelo ainda pode ser vista como uma área relevante de estudo. É fundamental a utilização de estratégias que auxiliem na especificação de efeitos fixos, aleatórios e de estruturas de covariâncias, bem como o uso de ferramentas que auxiliem na visualização das estruturas experimentais mais complexas.

Nos experimentos de adubação, por exemplo, é comum a utilização de esquemas fatoriais para avaliação de diferentes fontes e doses de fertilizantes ao longo do tempo. Outra prática comum refere-se à inclusão de um ou mais tratamentos adicionais ao fatorial a fim de realizar comparações, ou seja, servir como referência (padrão) para avaliação dos demais tratamentos envolvidos (YASSIN; MORAIS; MUNIZ, 2002). No entanto, a utilização dessa prática aumenta a complexidade das estruturas experimentais envolvidas, dificultando a formulação de modelos e, consequentemente, a realização das análises estatísticas.

Com base no exposto acima, o objetivo deste trabalho foi empregar a metodologia dos modelos lineares mistos no estudo da produtividade de matéria seca da parte aérea de plantas de alfafa a partir de cortes sucessivos, envolvendo ensaio em esquema fatorial com tratamento adicional, bem como propor a utilização do diagrama de Hasse para auxiliar na compreensão e visualização simbólica da estrutura experimental, obtenção dos números de graus de liberdade associados a cada fonte de variação e conjuntamente com a estratégia *top-down* facilitar a especificação do modelo linear misto.

2 REVISÃO BIBLIOGRÁFICA

2.1 Medidas repetidas

O termo medidas repetidas, de acordo com Diggle (1988), refere-se aos casos em que se observam uma ou mais variáveis respostas repetidamente na mesma unidade experimental. Segundo West, Welch e Galecki (2007), essas medidas podem ser coletadas sob diferentes condições experimentais ou observacionais como, por exemplo, em medidas de peso e tamanho corporal coletadas ao longo do tempo, mensurações de profundidade coletadas no espaço etc.

De acordo com Littell, Henry e Ammerman (1998), uma atenção especial deve ser dada à análise de dados com medidas repetidas devido a possíveis padrões de variação e correlação entre as respostas na mesma unidade amostral. Grande parte do esforço empregado na análise desses dados está relacionada à modelagem da estrutura de variâncias e covariâncias intra-unidades amostrais.

Segundo Littell, Henry e Ammerman (1998) e Wang e Goonewardene (2004), dentre outros, as abordagens comumente utilizadas na análise de dados com medidas repetidas estendem-se desde a utilização de modelos uni e multivariados até a metodologia baseada em modelos lineares mistos. As análises envolvendo o modelo univariado impõem uma restrição rigorosa à matriz de covariâncias. No modelo multivariado considera-se uma matriz de covariâncias sem restrições, isto é, não estruturada. Por sua vez, a abordagem por modelos mistos possibilita a utilização de diferentes estruturas para a matriz de variâncias e covariâncias.

Na abordagem mais simples, por meio do modelo univariado, supõe-se um delineamento em esquema de parcelas subdivididas para o experimento, no qual os tratamentos são aleatorizados às parcelas e os níveis do fator que se repete são considerados como os níveis de um suposto tratamento aplicado (sem aleatorização) às subparcelas. Supõe-se também a existência de duas fontes de variação residual, uma associada às parcelas e outra associada às subparcelas. Nessas condições, a estrutura da matriz de variâncias e covariâncias inerente ao modelo univariado no esquema de parcelas subdivididas, chamada de *uniforme* ou de *simetria composta*, é dada por:

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

em que σ_1^2 é a variância associada às parcelas e σ^2 é a variância associada às subparcelas dentro da mesma parcela.

Na estrutura de simetria composta, a variância entre as observações feitas na mesma parcela e mesmo nível do fator em que as medidas se repetem é igual a $(\sigma_1^2 + \sigma^2)$ e a covariância entre as observações feitas na mesma parcela e diferentes níveis do fator em que as medidas se repetem é igual a σ_1^2 . Isso pressupõe que as medidas possuem variâncias iguais ao longo de todo o processo de coleta e que os pares de medidas na mesma unidade experimental são igualmente correlacionados. Porém, segundo Littell, Henry e Ammerman (1998), medidas realizadas na mesma unidade experimental podem ser correlacionadas porque elas contém uma contribuição comum da unidade. Além disso, medidas realizadas em unidades próximas no tempo, por exemplo, tendem a ser mais correlacionadas do que medidas mais distantes e a variância tende a mudar ao longo do processo de coleta dessas medidas.

De acordo com Littell, Henry e Ammerman (1998), a validade dos erros padrões para estimativas de efeitos fixos pode ser comprometida ao se ignorar a dependência das medidas tomadas na mesma unidade experimental. Segundo Wang e Goonewardene (2004), o uso de erros padrões incorretos para comparações de médias pode resultar numa maior probabilidade de rejeitar a hipótese de que não existe diferença entre os tratamentos, quando esta é verdadeira (erro Tipo I).

Pelos motivos citados anteriormente, a análise univariada considerando o esquema de parcelas subdivididas nem sempre é recomendada para análise de dados com medidas repetidas. Essa análise só pode ser realizada se a condição proposta por Huynh e Feldt (1970) for satisfeita. A condição Huynh-Feldt (H-F) apresenta uma matriz de variâncias e covariâncias mais geral do que a matriz de simetria composta, de tal forma que as variâncias das diferenças entre pares de erros dessa matriz geral sejam todas iguais. No contexto de medidas repetidas, essa condição é necessária e suficiente para a validade do teste F da análise de variância com o esquema de parcelas subdivididas. O teste de *esfericidade* desenvolvido por Mauchly (1940) pode ser utilizado para verificar se a matriz de covariâncias atende à condição H-F. Esse teste avalia se uma população multivariada apresenta variâncias iguais e correlações nulas.

Uma abordagem tradicional, se o teste de esfericidade indicar que a matriz de covariâncias não atende à condição H-F, é a realização de análise univariada aproximada com correções dos números de graus de liberdade como aquelas propostas por Geisser e Greenhouse (1958) ou Huynh e Feldt (1976), porém essas correções nem sempre são adequadas (SAS INSTITUTE INC., 2002). Outra abordagem tradicional é a realização de uma análise de variância multivariada, também conhecida como análise multivariada de perfis, que requer a estimação de todos os parâmetros para a matriz de covariâncias, o que é muito mais geral do que necessário na maioria dos estudos envolvendo medidas repetidas. Assim, perde-se uma grande quantia de informações inerentes a esses dados, podendo resultar em testes menos poderosos (SAS INSTITUTE INC., 2002; WANG; GOONEWARDENE, 2004). Além disso, a análise multivariada de perfis depende de dados completos e balanceados. Por exemplo, se uma unidade experimental não possui dados para a variável resposta ao longo de todo o tempo, então essa unidade experimental é completamente omitida na análise.

Geralmente, não é possível, por parte do pesquisador, controlar todas as circunstâncias sobre as quais as medidas repetidas são coletadas. Isso gera conjuntos de dados incompletos e desbalanceados, que são difíceis de ser analisados usando-se o modelo multivariado geral com estrutura de covariância irrestrita ou não estruturada (LAIRD; WARE, 1982).

Por ser capaz de solucionar muitos dos problemas descritos anteriormente, a abordagem envolvendo a especificação de modelos lineares mistos em estudos com medidas repetidas é mais abrangente, precisa e recomendada em relação às abordagens tradicionais (WANG; GOONEWARDENE, 2004). Com modelos de efeitos mistos, é possível utilizar várias estruturas de covariâncias no processo de modelagem, oferecendo maior eficiência do que estimar todos os parâmetros para a matriz de covariâncias como realizado na análise de variância multivariada e maior flexibilidade em relação aos modelos univariados que satisfazem o teste de esfericidade (WEST; WELCH; GALECKI, 2007). Além disso, a maior flexibilidade na escolha da estrutura de covariâncias permite explicar adequadamente padrões de relação entre as respostas coletadas na mesma unidade experimental e fornece erros padrões válidos para estimativas de efeitos fixos (LITTELL; HENRY; AMMERMAN, 1998; WEST; WELCH; GALECKI, 2007).

Com relação à perda de observações, Cnaan, Laird e Slasor (1997) apresentam uma ampla discussão sobre modelos lineares mistos para dados de medidas repetidas e dados longitudinais considerando casos desbalanceados. Segundo Littell, Henry e Ammerman (1998), mesmo que uma unidade experimental não apresente todas as medidas, ela não será totalmente ignorada na análise por modelos mistos, ao contrário do que ocorre na análise de variância multivariada.

2.2 Modelos lineares mistos

Segundo Littell, Henry e Ammerman (1998), os modelos lineares mistos foram desenvolvidos por geneticistas para avaliar o potencial genético de touros. Sua aplicação foi disseminada por várias áreas de pesquisas científicas, estimulada pela disponibilidade de avanços computacionais. Antes desses avanços as análises de modelos mistos eram realizadas adaptando-se métodos desenvolvidos para modelos de efeitos fixos, o que trazia limitações na aplicabilidade, uma vez que as estruturas de covariâncias não eram modeladas.

O nome "modelos lineares mistos" vem do fato desses modelos serem lineares nos parâmetros e apresentarem efeitos fixos, além da média geral, e efeitos aleatórios, além do erro. Os efeitos fixos são representados por parâmetros desconhecidos associados a covariáveis contínuas ou a níveis de fatores categóricos. São restritos aos elementos testados e a seus níveis. Por sua vez, os efeitos aleatórios são representados por variáveis aleatórias que seguem uma determinada distribuição de probabilidade e estão associados a fatores amostrados aleatoriamente de uma população (PINHEIRO; BATES, 2000; WEST; WELCH; GALECKI, 2007).

Com a inclusão de efeitos aleatórios aos fatores, o modelo misto pode acomodar covariâncias entre as observações, sendo uma opção recomendada para a análise de dados com medidas repetidas. Além disso, a introdução desses efeitos permite modelar a variação individual sem comprometer as estimativas dos efeitos fixos e pode melhorar consideravelmente a qualidade preditiva do modelo.

Baseando-se nas ideias introduzidas por Harville (1977), Laird e Ware (1982) descrevem uma classe flexível de modelos lineares mistos em dois estágios. Nesses modelos a distribuição de probabilidade para os vetores respostas de diferentes indivíduos pertencem a uma única família, mas os efeitos aleatórios variam entre os indivíduos de acordo com uma distribuição especificada no segundo estágio. Isso significa que em um estudo envolvendo modelos de regressão, por exemplo, os efeitos fixos do primeiro estágio são responsáveis pela obtenção de uma curva polinomial média e os efeitos aleatórios do segundo estágio são responsáveis pela inclusão de desvios individuais em relação à essa curva média.

O modelo linear misto para uma determinada unidade experimental i pode ser expresso por (LAIRD; WARE, 1982):

$$Y_{i} = X_{i}\beta + Z_{i}b_{i} + \varepsilon_{i}, \qquad (1)$$

com $b_{i} \sim N_{q}(\mathbf{0}, \mathbf{D})$, $\varepsilon_{i} \sim N_{n_{i}}(\mathbf{0}, \mathbf{R}_{i})$ e
 $b_{1}, \dots, b_{m}, \varepsilon_{1}, \dots, \varepsilon_{m}$ independentes,

em que, \mathbf{Y}_i representa um vetor n_i -dimensional de variáveis respostas pertencentes à unidade experimental $i, 1 \leq i \leq m, m$ é o número de unidades experimentais; $\boldsymbol{\beta}$ é um vetor p-dimensional de parâmetros (efeitos fixos); \boldsymbol{b}_i é um vetor q-dimensional de efeitos aleatórios (refletem o comportamento individual da *i*-ésima unidade experimental) e segue uma distribuição normal q-variada com vetor de média $\mathbf{0}$ e matriz de covariâncias \boldsymbol{D} ; \boldsymbol{X}_i (de dimensão $n_i \times p$) e \boldsymbol{Z}_i (de dimensão $n_i \times q$) são matrizes de especificações conhecidas e de posto completo dos efeitos fixos e aleatórios, respectivamente; $\boldsymbol{\varepsilon}_i$ é um vetor n_i dimensional de erros e segue uma distribuição normal n_i -variada com vetor de média $\mathbf{0}$ e matriz de covariâncias \boldsymbol{R}_i .

Seguindo a notação de West, Welch e Galecki (2007), os elementos presentes em cada vetor ou matriz do modelo (1) podem ser representados a fim de facilitar a compreensão. Por exemplo, o vetor Y_i com as n_i respostas para *i*-ésima unidade experimental é conhecido como perfil individual de respostas em estudos longitudinais e pode ser expresso como:

$$oldsymbol{Y_i} = \left[egin{array}{c} Y_{1i} \ Y_{2i} \ dots \ Y_{n_ii} \end{array}
ight]$$

Deve-se ressaltar que o número de observações coletadas na *i*-ésima unidade experimental (n_i) presentes no vetor Y_i pode ser diferente de uma unidade experimental para outra.

A matriz X_i (de dimensão $n_i \times p$) no modelo (1) representa os valores conhecidos das p covariáveis, $X^{(1)}, \ldots, X^{(p)}$, para cada uma das n_i observações coletadas na *i*-ésima unidade experimental:

$$\boldsymbol{X_i} = \begin{bmatrix} X_{1i}^{(1)} & X_{1i}^{(2)} & \dots & X_{1i}^{(p)} \\ X_{2i}^{(1)} & X_{2i}^{(2)} & \dots & X_{2i}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_ii}^{(1)} & X_{n_ii}^{(2)} & \dots & X_{n_ii}^{(p)} \end{bmatrix}$$

O vetor β no modelo (1) é um vetor *p*-dimensional de coeficientes de regressão (ou parâmetros de *efeitos fixos*) associados as *p* covariáveis usadas na construção da matriz X_i :

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

A matriz Z_i (de dimensão $n_i \times q$) no modelo (1) representa os valores conhecidos das q covariáveis $Z^{(1)}, \ldots, Z^{(q)}$, para cada uma das n_i observações coletadas na *i*-ésima unidade experimental:

$$\boldsymbol{Z_i} = \begin{bmatrix} Z_{1i}^{(1)} & Z_{1i}^{(2)} & \dots & Z_{1i}^{(q)} \\ Z_{2i}^{(1)} & Z_{2i}^{(2)} & \dots & Z_{2i}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{nii}^{(1)} & Z_{nii}^{(2)} & \dots & Z_{nii}^{(q)} \end{bmatrix}$$

No modelo (1), b_i é um vetor q-dimensional de *efeitos aleatórios* associado

às q covariáveis da matriz Z_i , que pode ser representado por:

$$oldsymbol{b_i} = \left[egin{array}{c} b_{1i} \ b_{2i} \ dots \ b_{qi} \ dots \ b_{qi} \end{array}
ight]$$

O vetor $\boldsymbol{\varepsilon}_i$ no modelo (1) é um vetor n_i -dimensional de erros. Cada elemento em $\boldsymbol{\varepsilon}_i$ denota o erro associado a cada resposta observada na *i*-ésima unidade experimental. O número n_i de observações coletadas pode ser diferente entre as unidades experimentais, o que implica que o número de elementos de $\boldsymbol{\varepsilon}_i$ pode ser diferente para cada *i*. O vetor $\boldsymbol{\varepsilon}_i$ pode ser representado por:

$$oldsymbol{arepsilon_{i}} oldsymbol{arepsilon_{i}} = \left[egin{array}{c} arepsilon_{1i} \ arepsilon_{2i} \ arepsilon \ arepsilon_{nii} \end{array}
ight]$$

Na primeira suposição do modelo (1), $\boldsymbol{b}_i \sim N_q(\boldsymbol{0}, \boldsymbol{D})$, os elementos da diagonal principal da matriz \boldsymbol{D} representam a variância de cada efeito aleatório de \boldsymbol{b}_i e os elementos fora dessa diagonal representam as covariâncias entre os efeitos aleatórios correspondentes. Como cada unidade experimental *i* possui *q* efeitos aleatórios associados, tem-se que \boldsymbol{D} é uma matriz de dimensões $q \times q$ do tipo simétrica e positiva definida. Essa matriz poder ser representada da seguinte forma:

$$\boldsymbol{D} = Var(\boldsymbol{b}_i) = \begin{bmatrix} Var(b_{1i}) & cov(b_{1i}, b_{2i}) & \dots & cov(b_{1i}, b_{qi}) \\ cov(b_{1i}, b_{2i}) & Var(b_{2i}) & \dots & cov(b_{2i}, b_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(b_{1i}, b_{qi}) & cov(b_{2i}, b_{qi}) & \dots & Var(b_{qi}) \end{bmatrix}$$

Os elementos (covariâncias) da matriz D geralmente podem ser definidos como funções de um conjunto menor de parâmetros de covariâncias. Existem estruturas pré-definidas, que evidenciam a existência ou inexistência de relacionamento entre esses parâmetros. Na subseção (2.2.1) serão apresentadas algumas dessas estruturas. Na segunda suposição do modelo (1), $\boldsymbol{\varepsilon}_{i} \sim N_{n_{i}}(\mathbf{0}, \mathbf{R}_{i})$, os elementos da diagonal principal da matriz \mathbf{R}_{i} representam a variância de cada erro de $\boldsymbol{\varepsilon}_{i}$ e os elementos fora da diagonal principal representam as covariâncias entre os erros correspondentes. De modo geral, \mathbf{R}_{i} pode ser representada da seguinte forma:

$$\boldsymbol{R}_{i} = Var(\boldsymbol{\varepsilon}_{i}) = \begin{bmatrix} Var(\varepsilon_{1i}) & cov(\varepsilon_{1i}, \varepsilon_{2i}) & \dots & cov(\varepsilon_{1i}, \varepsilon_{nii}) \\ cov(\varepsilon_{1i}, \varepsilon_{2i}) & Var(\varepsilon_{2i}) & \dots & cov(\varepsilon_{2i}, \varepsilon_{nii}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\varepsilon_{1i}, \varepsilon_{nii}) & cov(\varepsilon_{2i}, \varepsilon_{nii}) & \dots & Var(\varepsilon_{nii}) \end{bmatrix}$$

Diferentemente da pressuposição $\mathbf{R}_i = \sigma^2 \mathbf{I}$ do modelo linear clássico, os erros associados às medidas repetidas realizadas na mesma unidade experimental em um modelo linear misto podem ser correlacionados. Isso implica que a matriz \mathbf{R}_i também pode assumir diferentes estruturas e algumas delas serão vistas na subseção (2.2.1).

O modelo (1) considera matricialmente o modelo linear misto para uma dada unidade experimental *i*. Segundo West, Welch e Galecki (2007), alternativamente, o modelo matricial pode ser extendido de tal modo que considere todas a unidades experimentais:

$$Y = X\beta + Zb + \varepsilon, \tag{2}$$

em que, \mathbf{Y} é um vetor de dimensão $n \times 1$, $n = \sum_{i=1}^{m} n_i$, construído por meio do empilhamento dos vetores \mathbf{Y}_i referentes a cada unidade experimental; \mathbf{X} é uma matriz de dimensão $n \times p$ obtida pelo empilhamento das matrizes \mathbf{X}_i ; \mathbf{Z} é uma matriz bloco diagonal em que os blocos são definidos pelas matrizes \mathbf{Z}_i ; os vetores $\mathbf{b} \in \boldsymbol{\varepsilon}$ são supostos independentes com $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}) \in \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$ e são construídos empilhando-se os vetores $\mathbf{b}_i \in \boldsymbol{\varepsilon}_i$, respectivamente; a matriz \mathbf{G} é uma matriz bloco diagonal, que representa a matriz de covariâncias para todos os efeitos aleatórios, seus blocos são definidos pela matriz \mathbf{D} ; a matriz \mathbf{R} é também uma matriz bloco diagonal, mas representando a matriz de covariâncias para todos os erros, e seus blocos são formados pelas matrizes \mathbf{R}_i . Assim, o modelo (2) pode ser melhor visualizado da seguinte forma:

$$\left[egin{array}{c} Y_1 \ Y_2 \ dots \ Y_m \end{array}
ight] = \left[egin{array}{c} X_1 \ X_2 \ dots \ X_m \end{array}
ight]eta + \left[egin{array}{c} Z_1 & 0 & \cdots & 0 \ 0 & Z_2 & \cdots & 0 \ dots \ Z_2 & \cdots & 0 \ dots & dots & \ddots & dots \ b_m \end{array}
ight] \left[egin{array}{c} b_1 \ b_2 \ dots \ dots$$

E as matrizes \boldsymbol{G} e \boldsymbol{R} :

$$G = \begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D \end{bmatrix} \quad e \quad R = \begin{bmatrix} R_1 & 0 & \cdots & 0 \\ 0 & R_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_m \end{bmatrix}$$

Considerando a suposição de independência entre \boldsymbol{b} e $\boldsymbol{\varepsilon}$ obtém-se a média e a matriz de covariâncias marginais de \boldsymbol{Y} :

$$E (\mathbf{Y}|\mathbf{b}) = E (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

$$Var (\mathbf{Y}|\mathbf{b}) = Var (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}) = \mathbf{R}$$

$$E (\mathbf{Y}) = E (E (\mathbf{Y}|\mathbf{b})) = E (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{B}) = \mathbf{X}\boldsymbol{\beta}$$

$$Var (\mathbf{Y}) = Var (E (\mathbf{Y}|\mathbf{b})) + E (Var (\mathbf{Y}|\mathbf{b})) = Var (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{B}) + E (\mathbf{R})$$

$$= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \mathbf{V}$$

Portanto, sob o modelo marginal, o vetor de respostas Y tem distribuição normal multivariada com vetor de médias $X\beta$ e matriz de covariâncias V = ZGZ' + R:

$$\boldsymbol{Y} \sim N\left(\boldsymbol{X}\boldsymbol{\beta}, \ \boldsymbol{V}\right)$$
 (3)

2.2.1 Estruturas das matrizes de covariâncias

As matrizes de covariâncias $D \in \mathbf{R}_i$ podem assumir diferentes estruturas. Algumas delas serão apresentadas a seguir para um caso particular, em que as matrizes possuem dimensão 4 × 4, ou seja, se for a matriz D, então o modelo linear misto possui quatro efeitos aleatórios (q = 4) associados com a *i*-ésima unidade experimental; se for a matriz \mathbf{R}_i , então foram coletadas quatro observações $(n_i = 4)$ em cada unidade experimental. O número de parâmetros para o caso geral t = q ou $t = n_i$ será considerado para cada uma das matrizes de covariâncias apresentadas.

1. Componentes de Variância ("Variance Components - VC")

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \boldsymbol{I}\sigma^2,$$

em que I é uma matriz identidade de dimensão 4×4 e σ^2 um parâmetro de variância. Essa estrutura matricial supõe independência e homogeneidade de variâncias entre os componentes, suposição utilizada nos modelos lineares clássicos de análise de variância e envolve um único parâmetro (σ^2).

2. Simetria Composta ("Compound Symmetry - CS")

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

A estrutura de simetria composta especifica em uma mesma unidade experimental covariâncias homogêneas (σ_1^2) e variâncias homogêneas ($\sigma_1^2 + \sigma^2$). Essa estrutura envolve dois parâmetros.

3. Simetria Composta com Heterogêneidade de Variâncias ("Heterogeneous CS - CSH")

$$\begin{array}{cccc} \sigma_1^2 & \sigma_1 \sigma_2 \rho & \sigma_1 \sigma_3 \rho & \sigma_1 \sigma_4 \rho \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_2 \sigma_3 \rho & \sigma_2 \sigma_4 \rho \\ \sigma_3 \sigma_1 \rho & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_3 \sigma_4 \rho \\ \sigma_4 \sigma_1 \rho & \sigma_4 \sigma_2 \rho & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{array}$$

4. Auto-Regressiva de Primeira Ordem ("First-Order Autoregressive - AR(1)")

$$\sigma^{2} \begin{bmatrix} 1 & \rho & \rho^{2} & \rho^{3} \\ \rho & 1 & \rho & \rho^{2} \\ \rho^{2} & \rho & 1 & \rho \\ \rho^{3} & \rho^{2} & \rho & 1 \end{bmatrix}$$

A estrutura auto-regressiva de primeira ordem, denotada por AR(1), especifica variâncias homogêneas (σ^2) e covariâncias diferentes, que tendem a zero à medida que se aumenta a distância entre as observações repetidas, uma vez que a covariância entre as observações é dada por uma função exponencial, envolvendo o parâmetro ρ , $-1 \leq \rho \leq 1$. Segundo West, Welch e Galecki (2007), a estrutura AR(1) é comumente usada como estrutura para a matriz ε_i no ajuste de modelos para dados com observações igualmente espaçadas na mesma unidade amostral. Nessa estrutura, as covariâncias entre observações próximas são maiores do que as covariâncias distantes no tempo ou no espaço, por exemplo, e envolve dois parâmetros.

 Auto-Regressiva de Primeira Ordem com Heterogêneidade de Variâncias ("Heterogeneous AR(1) - ARH(1)")

$$\begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho & \sigma_1 \sigma_3 \rho^2 & \sigma_1 \sigma_4 \rho^3 \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_2 \sigma_3 \rho & \sigma_2 \sigma_4 \rho^2 \\ \sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_3 \sigma_4 \rho \\ \sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{bmatrix}$$

É uma generalização da matriz AR(1), admitindo variâncias e covariâncias diferentes e diminuição da correlação. No caso geral, essa estrutura envolve t + 1 parâmetros.

 Auto-regressiva e Médias Móveis de Primeira Ordem ("First-Order Autoregressive Moving-Average - ARMA(1,1)")

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

A estrutura auto-regressiva e médias móveis de primeira ordem, denotada por ARMA(1,1), é uma estrutura associada a séries temporais com três parâmetros, o auto-regressivo ρ , componente de médias móveis γ e a variância residual σ^2 . A família completa ARMA(p,q), descrita em detalhes por Box, Jenkins e Reinsel (1994), forma uma classe de estruturas de correlação muito úteis e parcimoniosas para descrever dados de séries temporais, sendo alternativas naturais à estrutura completamente parametrizada.

7. Não estruturada ou completamente parametrizada ("Unstructured - UN")

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Esta matriz não apresenta padrão entre as covariâncias ($\sigma_{k,l}, k \neq l$) e nem sobre as variâncias (σ_k^2), sendo completamente geral. Por isso recebe o nome de "*não* estruturada", já que não existe condição estrutural matemática nas variâncias e covariâncias. Segundo Pinheiro e Bates (2000), se existirem poucas observações por unidade experimental, a estrutura de correlação geral é útil como uma ferramenta exploratória para determinar estruturas mais simples, que conduzam a modelos mais parcimoniosos. No caso geral, essa estrutura envolve t(t + 1)/2 parâmetros. 8. Componentes de Variância com Heterogeneidade ("Banded Main Diagonal - UN(1)")

$$\left[\begin{array}{ccccc} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{array}\right]$$

Esta matriz conserva a suposição de independência, mas não a de homogeneidade de variâncias entre os componentes ($\sigma_k^2 \neq \sigma_{k'}^2$, para $k \neq k'$). Segundo West, Welch e Galecki (2007), a estrutura simples com heterogeneidade (também chamada de matriz diagonal) é frequentemente utilizada para especificar a matriz de covariâncias dos efeitos aleatórios (D), em que todas as covariâncias entre esses efeitos são nulas, contudo cada efeito aleatório de b_i tem sua própria variância. No caso geral, essa estrutura envolve t parâmetros.

2.2.2 Especificações de estruturas de covariâncias

A fim de ilustrar a especificação de estruturas de covariâncias no ajuste de modelos lineares mistos, consideremos o exemplo de um experimento farmacêutico utilizado por Littell, Pendergast e Natarajan (2000). Nesse experimento, o objetivo do estudo foi comparar o efeito de dois medicamentos (A e B) e um placebo (P) em medidas de capacidade respiratória, chamadas de FEV1. Considera-se um delineamento completamente aleatorizado em que vinte e quatro pacientes foram atribuídos aleatoriamente a cada um dos três grupos de tratamentos (A,B e P) e FEV1 foi medida no início do estudo, antes da administração dos medicamentos, sendo chamada BASEFEV1, e depois em intervalos de uma hora, ao longo de oito horas.

Visando uma melhor exposição do conteúdo, segue uma descrição semelhante ao desenvolvimento teórico do modelo linear misto adotado pelos autores no trabalho em questão. A fim de ilustrar as construções das estruturas de covariâncias, sem se preocupar com maiores complicações, opta-se por um conjunto de dados balanceados. Além disso, assume-se que as medidas são independentes entre pacientes distintos.

Denota-se por Y_{ijk} a variável resposta no tempo k, no paciente j e no grupo i, i = 1, 2, 3, j = 1, 2, ..., 24, k = 1, 2, ..., 8. O valor esperado de Y_{ijk} refere-se à parte de efeitos fixos do modelo linear misto e é dado por $E(Y_{ijk}) = \mu_{ijk}$. O valor esperado, μ_{ijk} , geralmente é modelado como uma função do tratamento, tempo e outras covariáveis de efeitos fixos. Assume-se que os efeitos aleatórios são normalmente distribuídos e que a parte desses efeitos no modelo especifica a estrutura de covariância das observações. Pela suposição de independência entre os pacientes tem-se que $cov(Y_{ijk}, Y_{i'j'k}) = 0$ se $i \neq i'$ ou $j \neq j'$. Assume-se também que as variâncias e covariâncias das medidas em um único paciente são as mesmas dentro dos grupos. Porém, permite-se a possibilidade das variâncias não serem homogêneas em todos os tempos e a possibilidade das covariâncias entre observações em tempos diferentes no mesmo paciente não serem as mesmas em todos os pares de tempos.

Uma estrutura de covariância geral (não estruturada) é denotada por $cov(Y_{ijk}, Y_{ijl}) = \sigma_{k,l}$, em que $\sigma_{k,l}$ é a covariância entre medidas dos tempos k e l no mesmo paciente e $\sigma_{k,k} = \sigma_k^2$ denota a variância no tempo k, k e l = 1, 2, ..., 8.

Denota-se por $\mathbf{Y}_{ij} = [Y_{ij1}, Y_{ij2}, \dots, Y_{ij8}]'$ o vetor de dados do grupo *i*, paciente *j* e nos tempos 1, 2, ..., 8. Deste modo, em notação matricial, o modelo é escrito por:

$$\boldsymbol{Y}_{ij} = \boldsymbol{\mu}_{ij} + \boldsymbol{\varepsilon}_{ij}^* \tag{4}$$

em que $\boldsymbol{\mu}_{ij} = [\mu_{ij1}, \mu_{ij2}, \dots, \mu_{ij8}]'$ é o vetor de médias e $\boldsymbol{\varepsilon}_{ij} = [\varepsilon_{ij1}^*, \varepsilon_{ij2}^*, \dots, \varepsilon_{ij8}^*]'$ é o vetor de erros, respectivamente, para o paciente *j* no grupo *i*. Representações matriciais da esperança e variância de \boldsymbol{Y}_{ij} são $E(\boldsymbol{Y}_{ij}) = \boldsymbol{\mu}_{ij}$ e $V(\boldsymbol{Y}_{ij}) = \boldsymbol{V}_{ij}$, em que \boldsymbol{V}_{ij} é uma matriz $8 \times 8 \operatorname{com} \sigma_{k,l}$ na linha *k* e na coluna *l*.

$$\boldsymbol{V}_{ij} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \cdots & \sigma_{18} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \cdots & \sigma_{28} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \cdots & \sigma_{38} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \cdots & \sigma_{48} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{81} & \sigma_{82} & \sigma_{83} & \sigma_{84} & \cdots & \sigma_8^2 \end{bmatrix}$$
(5)

Assume-se que V_{ij} é a mesma para todos os pacientes (isto é, para todo *i* e *j*), mas o subscrito *ij* é usado para enfatizar que a matriz de covariâncias (V_{ij}) refere-se a um único paciente.

Assim, como em (2) descrito na subseção 2.2, Littell, Pendergast

e Natarajan (2000) também propõem uma extensão na representação do modelo de tal forma que envolva todas as unidades experimentais. Denota-se $\boldsymbol{Y} = (\boldsymbol{Y}'_{1,1}, \dots, \boldsymbol{Y}'_{1,24}, \boldsymbol{Y}'_{2,1}, \dots, \boldsymbol{Y}'_{2,24}, \boldsymbol{Y}'_{3,1}, \dots, \boldsymbol{Y}'_{3,24})'$ o vetor dos dados de todos os sujeitos e similarmente para o vetor de valores esperados e de er-

ros como $E(\mathbf{Y}) = \boldsymbol{\mu} = (\boldsymbol{\mu}'_{1,1}, \dots, \boldsymbol{\mu}'_{1,24}, \boldsymbol{\mu}'_{2,1}, \dots, \boldsymbol{\mu}'_{2,24}, \boldsymbol{\mu}'_{3,1}, \dots, \boldsymbol{\mu}'_{3,24})'$ e $\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}^{*'}_{1,1}, \dots, \boldsymbol{\varepsilon}^{*'}_{1,24}, \boldsymbol{\varepsilon}^{*'}_{2,1}, \dots, \boldsymbol{\varepsilon}^{*'}_{2,24}, \boldsymbol{\varepsilon}^{*'}_{3,1}, \dots, \boldsymbol{\varepsilon}^{*'}_{3,24})'$. Assim, tem-se o modelo:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}^* \tag{6}$$

е

$$V(\boldsymbol{Y}) = \boldsymbol{V} = diag\{\boldsymbol{V}_{ij}\}$$

em que $diag\{V_{ij}\}$ refere-se a uma matriz bloco diagonal com V_{ij} em cada bloco.

O modelo linear misto univariado para as medidas repetidas de FEV1 é dado por:

$$Y_{ijk} = \mu + \lambda x_{ij} + \alpha_i + d_{ij} + \tau_k + (\alpha \tau)_{ik} + e_{ijk}$$

$$\tag{7}$$

em que μ é uma constante comum a todas observações, λ é um coeficiente da covariável x_{ij} =BASEFEV1 para o paciente j no grupo do medicamento i, α_i é o efeito fixo correspondente ao medicamento i, τ_k é o efeito fixo correspondente à hora k, $(\alpha \tau)_{ik}$ é o efeito fixo da interação do medicamento i com a hora k, d_{ij} é o efeito aleatório correspondente ao paciente j que recebeu o medicamento i e segue uma distribuição normal com média zero e variância σ_d^2 e, e_{ijk} é o efeito aleatório correspondente ao erro associado à resposta Y_{ijk} do paciente j, que recebeu o medicamento i na hora k e segue uma distribuição normal com média zero e variância σ_e^2 , independente de d_{ij} . Assim, tem-se:

$$E(Y_{ijk}) = \mu_{ijk} = \mu + \lambda x_{ij} + \alpha_i + \tau_k + (\alpha \tau)_{ik}$$
$$Var(Y_{ijk}) = \sigma_d^2 + \sigma_e^2$$
$$cov(Y_{ijk}, Y_{ijl}) = \sigma_d^2 + cov(e_{ijk}, e_{ijl})$$

Reescrevendo o modelo (7) em notação matricial, tem-se:
$$Y = X\beta + Zb + \varepsilon, \tag{8}$$

em que, \mathbf{Y} é um vetor com 576 linhas (24 pacientes × 8 horas × 3 medicamentos) com todos os dados de todos os pacientes; \mathbf{X} é uma matriz de especificação conhecida dos parâmetros de efeitos fixos $\mu, \lambda, \alpha_i, \tau_k$, $(\alpha \tau)_{ik}$ e de dimensão 576 × 37 (576 linhas = 24 pacientes × 8 horas × 3 medicamentos e 37 colunas = 1 + 1 + 3 + 8 + 3 × 8 níveis de cada fator fixo); $\boldsymbol{\beta}$ é o vetor de parâmetros (efeitos fixos) com 37 linhas; \mathbf{Z} é a matriz de especificação conhecida (contém apenas zeros e uns) dos efeitos aleatórios d_{ij} e de dimensão 576 × 72 (576 linhas = 24 pacientes × 8 horas × 3 medicamentos e 72 colunas = 24 pacientes × 3 medicamentos); \boldsymbol{b} é o vetor dos efeitos aleatórios d_{ij} com 72 linhas correspondendo aos vinte e quatro efeitos aleatórios de pacientes dentro de cada um dos três medicamentos; $\boldsymbol{\varepsilon}$ é um vetor dos erros e_{ijk} com 576 linhas. Deste modo, no modelo (6) tem-se que $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}^* = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$.

Assim, como descrito em (3) na subseção 2.2, assumindo-se a pressuposição de independência entre os vetores $\boldsymbol{b} \in \boldsymbol{\varepsilon}$, tem-se que:

$$Var(\boldsymbol{Y}) = \boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$$
(9)

Observa-se que a estrutura de covariância V dada em (9) está em função de G e R. De acordo com Littell, Pendergast e Natarajan (2000) e Singer, Nobre e Rocha (2012), em muitas aplicações com medidas repetidas, ZGZ' representa a variabilidade entre unidades amostrais e R representa a variabilidade intra-unidades amostrais.

Segundo Singer, Nobre e Rocha (2012) e Wang e Goonewardene (2004), grande parte do esforço e da dificuldade empregada na modelagem de dados com medidas repetidas concentra-se na especificação correta da estrutura de covariâncias da matriz Vdada em (9). Geralmente, a estrutura de V depende da maneira pela qual as obervações foram obtidas e do conhecimento sobre o mecanismo gerador dessas observações.

De acordo com Littell, Pendergast e Natarajan (2000), a estrutura de covariâncias da matriz V é construída em função da especificação das matrizes G e R. Assim como no trabalho desses autores, as notações das sub-matrizes de V, Z, G e Rcorrespondente ao paciente j dentro do medicamento i serão denotadas por $V_{i,j}$, $Z_{i,j}$, $G_{i,j}$ e $R_{i,j}$, respectivamente. Seguem as representações das matrizes e de suas respectivas sub-matrizes:

$$V_{576 \times 576} = Z_{576 \times 72} \quad G_{72 \times 72} \quad Z'_{72 \times 576} \quad + \quad R_{576 \times 576}$$

 $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{1,1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{1,2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{V}_{3,24} \end{bmatrix}, \text{ com } \boldsymbol{V}_{i,j} \text{ representado como em (5) para } i = 1, \dots, 3 \text{ e } j = 1, \dots, 24$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{1,1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{1,2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Z}_{3,24} \end{bmatrix}, \quad \text{com} \quad \boldsymbol{Z}_{i,j} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ para } i = 1, \dots, 3 \text{ e } j = 1, \dots, 24$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{D} \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & & \\ & & & & & \\ & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & \\ & & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & &$$

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{1,1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_{1,2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{R}_{3,24} \end{bmatrix}, \text{com } \boldsymbol{R}_{i,j} \text{ para } i = 1, \dots, 3 \text{ e } j = 1, \dots, 24 \text{ representado por }$$

$$\boldsymbol{R}_{i,j} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} & \sigma_{e_{13}} & \sigma_{e_{14}} & \cdots & \sigma_{e_{18}} \\ \sigma_{e_{21}} & \sigma_{e_{2}}^2 & \sigma_{e_{23}} & \sigma_{e_{24}} & \cdots & \sigma_{e_{28}} \\ \sigma_{e_{31}} & \sigma_{e_{32}} & \sigma_{e_{3}}^2 & \sigma_{e_{34}} & \cdots & \sigma_{e_{38}} \\ \sigma_{e_{41}} & \sigma_{e_{42}} & \sigma_{e_{43}} & \sigma_{e_{4}}^2 & \cdots & \sigma_{e_{48}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{e_{81}} & \sigma_{e_{82}} & \sigma_{e_{83}} & \sigma_{e_{84}} & \cdots & \sigma_{e_{88}}^2 \end{bmatrix}$$
(10)

De acordo com Littell, Pendergast e Natarajan (2000), considerando-se o

modelo 8, seguem algumas formas de especificar estruturas de covariâncias para a matriz V, especificamente em termos de $V_{i,j}$:

(i) Simples (S)

A estrutura de componentes de variância para a matriz V pode ser construída especificando G = 0 e $R = diag\{R_{i,j}\}$, em que $R_{i,j} = \sigma_S^2 I_{8\times8}$ e $I_{8\times8}$ é uma matriz identidade de dimensão (8 × 8). V = Z0Z' + R

$$V = Z 0 Z' + I$$

 $V = R$

Em termos de $V_{i,j}$ tem-se:

$$V_{i,j} = \begin{bmatrix} \sigma_S^2 & 0 & \cdots & 0 \\ 0 & \sigma_S^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_S^2 \end{bmatrix}$$

(ii) Simetria composta (SC)

Existem duas formas de especificar $\boldsymbol{G} \in \boldsymbol{R}$ a fim de se obter a estrutura simétrica composta para a matriz \boldsymbol{V} . A primeira consiste em definir $\boldsymbol{G} = \sigma_{SC,e}^2 \boldsymbol{I}_{72\times72}$ e $\boldsymbol{R} =$ $\sigma_{SC,d}^2 \boldsymbol{I}_{576\times576}$, em que os subscritos $d \in e$ identificam, respectivamente, a variação dentro e entre-pacientes. A segunda forma consiste em definir $\boldsymbol{G} = \boldsymbol{0} \in \boldsymbol{R}_{ij} =$ $\sigma_{SC,e}^2 \boldsymbol{J}_{8\times8} + \sigma_{SC,d}^2 \boldsymbol{I}_{8\times8}$, em que \boldsymbol{J} é uma matriz de uns. Nas duas formas, as submatrizes $\boldsymbol{V}_{i,j}$ resultantes serão iguais a:

$$\boldsymbol{V}_{i,j} = \begin{bmatrix} \sigma_{SC,e}^{2} + \sigma_{SC,d}^{2} & \sigma_{SC,e}^{2} & \cdots & \sigma_{SC,e}^{2} \\ \sigma_{SC,e}^{2} & \sigma_{SC,e}^{2} + \sigma_{SC,d}^{2} & \cdots & \sigma_{SC,e}^{2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{SC,e}^{2} & \sigma_{SC,e}^{2} & \cdots & \sigma_{SC,e}^{2} + \sigma_{SC,d}^{2} \end{bmatrix}$$

(iii) Auto-regressiva de primeira ordem - AR(1)

Essa estrutura é especificada somente em termos de \mathbf{R} , com $\mathbf{G} = \mathbf{0}$. Em $\mathbf{R}_{i,j}$, o elemento na linha k e coluna l (k e l = 1, 2, ..., 8) é definido como $\sigma_{AR(1)}^2 \rho^{|k-l|}$, envolvendo o parâmetro ρ , $-1 \leq \rho \leq 1$. Assim, tem-se que:

$$\boldsymbol{V}_{i,j} = \boldsymbol{R}_{i,j} = \sigma_{AR(1)}^{2} \begin{bmatrix} 1 & \rho & \rho^{2} & \rho^{3} & \cdots & \rho^{7} \\ \rho & 1 & \rho & \rho^{2} & \cdots & \rho^{6} \\ \rho^{2} & \rho & 1 & \rho & \cdots & \rho^{5} \\ \rho^{3} & \rho^{2} & \rho & 1 & \cdots & \rho^{4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{7} & \rho^{6} & \rho^{5} & \rho^{4} & \cdots & 1 \end{bmatrix}_{8 \times 8}$$

(iv) Não estruturada (NE)

A matriz não estruturada para V, especificamente para $V_{i,j}$, pode ser construída definindo G = 0 e uma sub-matriz $R_{i,j}$ de R completamente geral. Deste modo, tem-se que:

$$\boldsymbol{V}_{i,j} = \boldsymbol{R}_{i,j} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} & \sigma_{e_{13}} & \sigma_{e_{14}} & \cdots & \sigma_{e_{18}} \\ \sigma_{e_{21}} & \sigma_{e_2}^2 & \sigma_{e_{23}} & \sigma_{e_{24}} & \cdots & \sigma_{e_{28}} \\ \sigma_{e_{31}} & \sigma_{e_{32}} & \sigma_{e_3}^2 & \sigma_{e_{34}} & \cdots & \sigma_{e_{38}} \\ \sigma_{e_{41}} & \sigma_{e_{42}} & \sigma_{e_{43}} & \sigma_{e_{4}}^2 & \cdots & \sigma_{e_{48}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{e_{81}} & \sigma_{e_{82}} & \sigma_{e_{83}} & \sigma_{e_{84}} & \cdots & \sigma_{e_{8}}^2 \end{bmatrix}$$

2.2.3 Modelo linear misto marginal

O modelo linear misto (1) descrito na seção (2.2) implica no seguinte modelo linear marginal:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i^*, \tag{11}$$

em que $\boldsymbol{\varepsilon}_{i}^{*} \sim N(\mathbf{0}, \boldsymbol{V}_{i})$, ou seja, o vetor de resíduos $\boldsymbol{\varepsilon}_{i}^{*}$ segue uma distribuição normal com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\boldsymbol{V}_{i} = \boldsymbol{Z}_{i}\boldsymbol{D}\boldsymbol{Z}_{i}^{'} + \boldsymbol{R}_{i}$.

Segundo West, Welch e Galecki (2007), o modelo marginal (11) e o modelo modelo linear misto (1) descrito na seção (2.2) envolvem o mesmo número de parâmetros de covariâncias, porém as restrições impostas no espaço paramétrico no modelo linear misto são maiores do que no modelo marginal. De modo geral, as matrizes D e R_i em modelos lineares mistos têm de ser positivas definidas, em outras palavras, todos os seus autovalores devem ser estritamente positivos. Por sua vez, no modelo marginal a condição é que somente a matriz \boldsymbol{V}_i seja positiva definida.

O conceito de modelo marginal é importante, pois a estimação de efeitos fixos e parâmetros de covariâncias no modelo linear misto envolve um estágio que depende do modelo marginal. Além disso, nos casos em que os procedimentos computacionais gerem uma matriz que não seja positiva definida para D, condição inválida em um modelo linear misto, novas tentativas podem ser realizadas usando-se um modelo marginal implícito com um número menor de restrições (WEST; WELCH; GALECKI, 2007).

O vetor \boldsymbol{Y}_i , definido no modelo marginal (11), segue a seguinte distribuição marginal:

$$\boldsymbol{Y}_{i} \sim N_{n_{i}} \left(\boldsymbol{X}_{i} \boldsymbol{\beta}, \quad \boldsymbol{Z}_{i} \boldsymbol{D} \boldsymbol{Z}_{i}^{\prime} + \boldsymbol{R}_{i} \right)$$
 (12)

Segundo Verbeke e Molenbergs (2000), o modelo linear misto (1) descrito na seção (2.2) é definido por meio da distribuição $f(\boldsymbol{y}_i|\boldsymbol{b}_i)$ e $f(\boldsymbol{b}_i)$, que é chamada de formulação hierárquica do modelo linear misto. Os autores enfatizam que o modelo hierárquico (1) não é equivalente ao modelo marginal (11), uma vez que inferências baseadas no modelo marginal não assumem explicitamente a presença de efeitos aleatórios representando a heterogeneidade natural entre unidades experimentais no caso de dados longitudinais.

2.2.4 Estimação em modelos lineares mistos

Para realizar inferências em modelos mistos é preciso estimar os parâmetros desconhecidos de β , D e R_i e predizer os efeitos aleatórios de b_i . São dois os casos de estimação a serem considerados seguindo a suposição de normalidade dos vetores Y e b_i . No primeiro caso, tomam-se como conhecidos os parâmetros de D e R_i e obtémse soluções para β e b_i . No segundo caso, o qual é o mais comumente encontrado, os parâmetros de D e R_i são desconhecidos (LITTELL et al., 2006).

Segundo Pinheiro e Bates (2000), dentre os vários métodos de estimação para modelos lineares mistos, os mais comumente utilizados são os métodos da Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR). De acordo com Littell et al. (2006), embora existam vários métodos para estimação dos parâmetros de covariâncias (Métodos dos Momentos, Método da Estimação Quadrática Não-viesada de Variância Mínima, Máxima Verossimilhança, Máxima Verossimilhança Restrita e Pseudo-Verossimilhança), o Método da Máxima Verossimilhança Restrita é indiscutivelmente o mais importante dentre eles. Os detalhes da estimação por MV e MVR são dados nas próximas subseções.

2.2.4.1 Máxima Verossimilhança - MV

O método de estimação por máxima verossimilhança (MV) consiste basicamente no processo de maximização conjunta da função dos parâmetros de uma distribuição conhecida, denominada função de verossimilhança.

Segundo West, Welch e Galecki (2007), no contexto de modelos lineares mistos, constrói-se a função de verossimilhança de β e θ (vetor com os parâmetros de De \mathbf{R}_i) com base na distribuição marginal da variável dependente \mathbf{Y}_i definida em (12) da seção 2.2.3. A função densidade de probabilidade normal multivariada correspondente é:

$$f(\boldsymbol{Y}_{i}|\boldsymbol{\beta},\boldsymbol{\theta}) = (2\pi)^{-\frac{n_{i}}{2}} |\boldsymbol{V}_{i}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{Y}_{i}-\boldsymbol{X}_{i}\boldsymbol{\beta})'\boldsymbol{V}_{i}^{-1}(\boldsymbol{Y}_{i}-\boldsymbol{X}_{i}\boldsymbol{\beta})\right]$$
(13)

em que $\boldsymbol{\theta}$ é o conjunto de parâmetros associados aos componentes da variância do modelo.

Com base na função densidade de probabilidade definida em (13) e considerando-se os dados observados, $Y_i = y_i$, a função de verossimilhança para a *i*-ésima unidade experimental é definida como:

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_i) = (2\pi)^{-\frac{n_i}{2}} |\boldsymbol{V}_i|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})\right]$$
(14)

A função de verossimilhança, $L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})$, é definida como produto das m contribuições independentes definidas na equação (14):

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{m} L_{i}(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}_{i})$$

$$= \prod_{i=1}^{m} (2\pi)^{-\frac{n_{i}}{2}} |\boldsymbol{V}_{i}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{y}_{i} - \boldsymbol{X}_{i}\boldsymbol{\beta})'\boldsymbol{V}_{i}^{-1}(\boldsymbol{y}_{i} - \boldsymbol{X}_{i}\boldsymbol{\beta})\right]$$

Assim, o logaritmo de $L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})$ é dado por

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = ln [L(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})]$$

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y}) = -\frac{n}{2} ln(2\pi) - \frac{1}{2} \sum_{i=1}^{m} ln(|\boldsymbol{V}_i|) - \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}),$$
(15)

em que $n = \sum_{i=1}^{m} n_i$ é o número de observações (linhas) no conjunto de dados.

O logaritmo da função de verossimilhança depende tanto dos parâmetros contidos em β , quanto dos componentes de variância de V_i contidos em θ . O problema consiste na maximização da função de verossimilhança para obtenção das estimativas de β e θ . Entretanto, muitos algoritmos computacionais simplificam o processo de otimização por meio da obtenção da função de verossimilhança perfilada como mostrado a seguir:

(i) Caso especial: assumindo-se θ conhecido

Considerando conhecidos os componentes de variância do modelo, ou seja, $\boldsymbol{\theta}$ conhecido, os únicos parâmetros a serem estimados são os efeitos fixos ($\boldsymbol{\beta}$). Portanto, o logaritmo da função de verossimilhança, $l(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})$, está apenas em função de $\boldsymbol{\beta}$ e sua otimização equivale a encontrar o mínimo da função $q(\boldsymbol{\beta})$ definida pelo último termo da função (15):

$$q(\boldsymbol{\beta}) = \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})' \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$$
(16)

Note a semelhança da função acima, exceto pela ponderação da matriz V_i^{-1} , com a forma matricial para a soma de quadrados que é minimizada no processo de estimação em um modelo linear clássico. Assim, pelo método dos mínimos quadrados generalizados, a solução para β , condicionada a θ , é dada por

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \boldsymbol{V}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \boldsymbol{V}_{i}^{-1} \boldsymbol{y}_{i}, \qquad (17)$$

tal que,

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{'} \boldsymbol{V}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1}\right)$$

Por apresentar propriedades estatísticas desejáveis, o estimador $\hat{\boldsymbol{\beta}}$ recebe o nome de BLUE (*Best Linear Unbiased Estimator* - melhor estimador linear não viesado). "Melhor" no sentido que minimiza a variância amostral, "linear", porque são funções lineares de \boldsymbol{y} e "não viesados", pois $E(\boldsymbol{\beta}) = \boldsymbol{\beta}$.

A solução $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ foi escrita condicionada a $\boldsymbol{\theta}$, a fim de tornar explícita a dependência dos parâmetros de covariância. O relacionamento entre os parâmetros de covariância, $\boldsymbol{\theta}$, e o valor de $\boldsymbol{\beta}$ que maximiza $l(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})$ será utilizado no próximo item para se obter a função de verossimilhança perfilada de modo que seu logaritmo fique estritamente em função de $\boldsymbol{\theta}$.

(ii) Caso geral: assumindo-se θ desconhecido

Neste caso, é preciso estimar os componentes de variância de $\boldsymbol{\theta}$ por meio do logaritmo da função de verossimilhança perfilada $l_{MV}(\boldsymbol{\theta}; \boldsymbol{y})$. Isso é feito substituindo $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ encontrado em (17) na função $l(\boldsymbol{\beta}, \boldsymbol{\theta}; \boldsymbol{y})$ descrita em (15). Os únicos componentes desconhecidos de

$$l_{MV}(\boldsymbol{\theta}; \boldsymbol{y}) = -\frac{n}{2} ln(2\pi) - \frac{1}{2} \sum_{i=1}^{m} ln(|\boldsymbol{V}_i|) - \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))' \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$$
(18)

são os parâmetros de covariância em $\boldsymbol{\theta}$.

Geralmente, a maximização de $l_{MV}(\boldsymbol{\theta}; \boldsymbol{y})$ em relação a $\boldsymbol{\theta}$ envolve processos computacionais iterativos, sendo um exemplo de otimização não linear, em que restrições são impostas a $\boldsymbol{\theta}$ de modo que as matrizes \boldsymbol{D} e \boldsymbol{R}_i sejam positiva definidas. Obtendo-se a estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ e consequentemente as matrizes $\hat{\boldsymbol{D}}$ e $\hat{\boldsymbol{R}}_i$, calcula-se uma estimativa para \boldsymbol{V}_i :

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_{i} = \boldsymbol{Z}_{i} \hat{\boldsymbol{D}} \boldsymbol{Z}_{i}^{'} + \hat{\boldsymbol{R}}_{i}$$
(19)

Desta forma, o estimador $\hat{\beta}$ dos parâmetros de efeitos fixos é obtido substituindo $\hat{V}(\hat{\theta})_i$ na equação (17):

$$\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_{i}^{-1} \boldsymbol{y}_{i}$$
(20)

A solução $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ foi escrita condicionada a $\hat{\boldsymbol{\theta}}$, a fim de tornar evidente a dependência da estimativa MV dos parâmetros de covariância na estimativa dos parâmetros de efeitos fixos.

O estimador $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$, por conter $\hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_i$ em sua formulação, é chamado de EBLUE (*Empirical Best Linear Unbiased Estimator*). Sua variância é dada por:

$$Var\left(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})\right) = \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1}$$
(21)

Por ignorar a variabilidade introduzida ao se trabalhar com as estimativas dos componentes de variância em vez dos valores paramétricos verdadeiros (desconhecidos), $Var\left(\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})\right)$ apresenta viés descendente.

Os estimadores MV de $\boldsymbol{\theta}$, geralmente, apresentam viés para pequenas amostras, pois não consideram a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo. De um modo alternativo, a correção do viés pode ser realizada utilizando-se o método da máxima verossimilhança restrita (MVR), que tende a corrigir alguns dos problemas da estimação por máxima verossimilhança.

2.2.4.2 Máxima Verossimilhança Restrita - MVR

A estimação pelo método da máxima verossimilhança restrita (MVR) é uma das técnicas mais importantes para estimação dos parâmetros de covariância em modelos lineares mistos. Os primeiros registros desse método foram realizados por Patterson e Thompson (1971) em um trabalho envolvendo delineamentos desbalanceados e blocos incompletos.

Em um modelo linear misto, os estimadores MVR de $\hat{\theta}$ são baseados na otimização da seguinte função:

$$l_{MVR}(\boldsymbol{\theta}; \boldsymbol{y}) = -\frac{(n-p)}{2} ln(2\pi) - \frac{1}{2} \sum_{i=1}^{m} ln(|\boldsymbol{V}_{i}|) - \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))' \boldsymbol{V}_{i}^{-1}(\boldsymbol{y}_{i} - \boldsymbol{X}_{i} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) - \frac{1}{2} \sum_{i=1}^{m} ln\left(|\boldsymbol{X}_{i}' \boldsymbol{V}_{i}^{-1} \boldsymbol{X}_{i}|\right)$$

$$(22)$$

em que p é o número de parâmetros de efeitos fixos.

De modo geral, não há uma forma explícita para a maximização de $l_{MVR}(\boldsymbol{\theta}; \boldsymbol{y})$ em relação a $\boldsymbol{\theta}$ para muitos modelos de efeitos mistos. As estimativas MVR são obtidas por métodos numéricos iterativos tal como o método de *Newton-Raphson*, o método *Fisher scoring* e o método EM proposto por Laird e Ware (1982). Nesses métodos, a otimização da função envolvida começa a partir de valores iniciais dos parâmetros de covariâncias ($\boldsymbol{\theta}^0$) e subsequentes iterações, controladas por um critério de convergência, são realizadas para encontrar os valores dos parâmetros que maximizam a função de verossimilhança.

Após obter a estimativa MVR de $\boldsymbol{\theta}$, a matriz $\hat{\boldsymbol{V}}_i$ pode ser calculada utilizando-se a mesma equação (19) apresentada no item (ii) da estimação MV e as estimativas dos parâmetros de efeitos fixos $\hat{\boldsymbol{\beta}}$ e sua variância $Var(\hat{\boldsymbol{\beta}})$ podem ser obtidas por meio das equações (20) e (21), respectivamente, descritas nesse mesmo item. Apesar da utilização das mesmas equações, os resultados para $\hat{\boldsymbol{\beta}} \in Var(\hat{\boldsymbol{\beta}})$ são diferentes na estimação MVR e MV, uma vez que a matriz $\hat{\boldsymbol{V}}_i$ é diferente em cada caso (WEST; WELCH; GALECKI, 2007).

A estimação MVR pode ser considerada como uma estimação MV para dados transformados, pois ao invés do logaritmo de Y, considera-se o logaritmo de $Y^* = LY$, em que L é uma matriz com n - posto(X) colunas (n é a dimensão de Y), de posto completo, com colunas ortogonais as colunas da matriz X, ou seja, LX = 0(ALCARDE, 2012; LITTELL et al., 2006). Isto significa que os estimadores dos componentes de variância são obtidos pela maximização da parte da função de verossimilhança que é invariante aos efeitos fixos do modelo. Por levar em consideração a perda dos graus de liberdade envolvidos na estimação dos parâmetros de efeitos fixos do modelo, as estimativas MVR de θ tendem a apresentar menor viés do que as estimativas MV (HARVILLE, 1977; WEST; WELCH; GALECKI, 2007).

2.2.5 Predição em modelos lineares mistos

Em modelos lineares mistos, além da estimação dos parâmetros de β , De \mathbf{R}_i também é possível predizer os efeitos aleatórios de \mathbf{b}_i , que refletem como os perfis individuais afastam-se do perfil médio geral. Esses efeitos não são parâmetros desconhecidos e sim variáveis aleatórias que seguem uma distribuição normal multivariada e por isso se diz "predizer", predizer valores dos efeitos aleatórios. A expressão para b_i pode ser derivada de uma extensão do teorema de Gauss-Markov sobre efeitos aleatórios (HARVILLE, 1976) ou por um enfoque empírico bayesiano baseado na média da distribuição posteriori. A expressão é dada por:

$$\hat{\boldsymbol{b}}_{i} = E(\boldsymbol{b}_{i}|\boldsymbol{Y}_{i} = \boldsymbol{y}_{i}) = \boldsymbol{D}\boldsymbol{Z}_{i}^{'}\boldsymbol{V}_{i}^{-1}(\boldsymbol{y}_{i} - \boldsymbol{X}_{i}\boldsymbol{\beta})$$
(23)

Geralmente, o parâmetro de efeitos fixos β é substituído por sua estimativa $\hat{\beta}$ e desta forma essa esperança condicional é conhecida como BLUP (*Best Linear Unbiased Predictors* - melhor preditor não viesado) dos efeitos aleatórios. Porém, como visto na equação (20) descrita na subseção 2.2.4.1, D, V_i e β estão condicionados a θ , que pode ser substituído por sua estimativa de máxima verossimilhança restrita $\hat{\theta}$. Assim,

$$\hat{\boldsymbol{b}}_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{D}} \boldsymbol{Z}_{\boldsymbol{i}}' \hat{\boldsymbol{V}}(\hat{\boldsymbol{\theta}})_{\boldsymbol{i}}^{-1} (\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}} \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}))$$
(24)

e por essa substituição recebe o nome de BLUP empírico (EBLUP).

A matriz de covariâncias dos EBLUPs é dada por (WEST; WELCH; GA-LECKI, 2007):

$$Var\left(\hat{\boldsymbol{b}}_{\boldsymbol{i}}\right) = \hat{\boldsymbol{D}}\boldsymbol{Z}_{\boldsymbol{i}}^{\prime} \left\{ \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1} - \hat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1}\boldsymbol{X}_{\boldsymbol{i}} \left(\sum_{i}^{m}\boldsymbol{X}_{i}\hat{\boldsymbol{V}}_{i}^{-1}\boldsymbol{X}_{i}\right)^{-1}\boldsymbol{X}_{i}\hat{\boldsymbol{V}}_{\boldsymbol{i}}^{-1} \right\} \boldsymbol{Z}_{\boldsymbol{i}}\hat{\boldsymbol{D}}$$
(25)

Deve-se ressaltar que (24) e (25) subestimam a verdadeira variabilidade na estimativa obtida \hat{b}_i por não levarem em conta a variabilidade introduzida pela substituição do parâmetro desconhecido θ por suas estimativas (VERBEKE; MOLENBERGHS, 2000).

2.2.6 Estimação e predição pelas equações de modelos mistos

As soluções de $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{b}}_i$ apresentadas anteriormente exigem o cálculo da matriz \boldsymbol{V}_i , que pode ser suficientemente grande em aplicações do melhoramento animal e vegetal, demandando alto poder computacional em processos de inversão de matrizes.

Sob a notação do modelo dado em (2) na subseção 2.2, Henderson (1950) apresentou as equações de modelos mistos para estimar β e **b** simultaneamente, sem a necessidade de calcular V^{-1} . Mais detalhes podem ser encontrados em Henderson(1963, 1984). Essas equações são obtidas pela maximização da função de verossimilhança conjunta de $\boldsymbol{Y} \in \boldsymbol{b}$ em relação aos vetores $\boldsymbol{\beta} \in \boldsymbol{b}$ e são dadas por:

$$\left[egin{array}{ccc} oldsymbol{X}^{'}oldsymbol{R}^{-1}oldsymbol{X} & oldsymbol{X}^{'}oldsymbol{R}^{-1}oldsymbol{Z} & \ oldsymbol{Z}^{'}oldsymbol{R}^{-1}oldsymbol{X} & oldsymbol{Z}^{'}oldsymbol{R}^{-1}oldsymbol{Z} & \ oldsymbol{eta} & \ oldsymbol{B} & \ oldsymbol{eta} & \ oldsymbol{B} & \ oldsymbol{A} & \ oldsymbol{B} & \ oldsymbol{B} & \ oldsymbol{B} & \ oldsymbol{A} & \ oldsymbol{B} & \ oldsymbol{A} & \ oldsymbol{B} & \ oldsymbol$$

As soluções do sistema podem ser obtidas por absorção ou por obtenção da matriz inversa por partição. Nos dois casos, os resultados serão:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$
(26)

е

$$\hat{\boldsymbol{b}} = \boldsymbol{D}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$
(27)

O estimador (26) é equivalente ao estimador (17), chamado melhor estimador linear não viesado de β , enquanto o preditor (27) é equivalente ao preditor (23), chamado melhor preditor linear não viesado de **b**.

Embora tenha-se referido inicialmente à maximização da função de verossimilhança conjunta, a função maximizada não é a verdadeira verossimilhança. A estimação e predição pelas equações de modelos mistos não devem ser confundidas com a estimação por máxima verossimilhança (restrita) dos parâmetros de covariância $\boldsymbol{\theta}$ e sim como apenas um método de estimação de $\boldsymbol{\beta}$ e \boldsymbol{b} (LITTELL et al., 2006).

2.2.7 Diagrama de Hasse na construção de modelos lineares mistos

A construção de modelos lineares mistos em delineamentos mais complexos nem sempre é algo trivial. Em estudos envolvendo dados longitudinais, por exemplo, além dos fatores longitudinais também podem existir outros fatores como os associados às unidades de observação como um resultado da aleatorização (fatores aleatorizados) e os fatores intrínsecos ao experimento (fatores não aleatorizados), que indexariam as unidades observacionais se nenhuma aleatorização tivesse sido realizada (BRIEN, 2015). Desta forma, a utilização de uma ferramenta visual capaz de organizar e estruturar os fatores em estudo torna-se relevante.

O diagrama de Hasse, assim chamado em homenagem ao matemático Helmut Hasse (1898-1979), é uma representação advinda da teoria dos conjuntos. O seu uso na área experimental iniciou-se com Taylor e Hilton (1981), destacando a sua utilização na representação da estrutura da análise de variância para delineamentos ortogonais, bem como para a obtenção do número de graus de liberdade, esperanças dos quadrados médios e de razões corretas para realização do teste F.

A construção do diagrama de Hasse para experimentos ortogonais envolve vários procedimentos, como descrição das características pertinentes ao estudo, determinação da estrutura experimental, representação de fatores generalizados, obtenção do números de graus de liberdade, obtenção das matrizes núcleo das formas quadráticas referentes às somas de quadrados e obtenção das esperanças dos quadrados médios. A descrição detalhada de cada uma dessas regras pode ser encontrada em Alcarde (2012) e Brien (2015).

Vale ressaltar que o diagrama de Hasse foi desenvolvido para experimentos ortogonais e a inclusão de um fator longitudinal quebra a ortogonalidade, modificando a forma de obtenção dos quadrados médios. Apesar disso, o diagrama de Hasse não deixa de ser uma ótima forma de visualizar simbolicamente o relacionamento entre os fatores experimentais (ALCARDE, 2012). Exemplos de utilização do diagrama de Hasse podem ser encontrados em Alcarde (2007, 2012), Brien (2015) e Lohr (1995).

2.2.8 Seleção de modelos

Diferentes modelos podem ser ajustados a conjuntos de dados a fim de identificar padrões consistentes e relações entre variáveis. Um modelo adequado é capaz de descrever fontes de variação nas variáveis dependentes ao mesmo tempo que permite testar hipóteses envolvidas na pesquisa. Além disso, é importante que esse modelo seja parcimonioso em relação ao número de parâmetros e que seja o melhor na predição da variável resposta estudada.

O processo de construção e seleção de modelos nem sempre é uma tarefa fácil, principalmente no caso de modelos lineares mistos pois além da seleção de efeitos fixos e identificação de efeitos aleatórios, também envolve a comparação de ajustes de estruturas de covariâncias.

De acordo com Diggle (1988), a estrutura de covariâncias de um modelo linear misto deve ser flexível, mas econômica. O autor ressalta que, embora esta estrutura não seja de interesse direto, uma superparametrização conduzirá a uma estimação ineficiente e potencialmente pobre na avaliação dos erros padrões das estimativas de perfis médios de resposta. Por outro lado, uma especificação restrita da estrutura de covariâncias, incapaz de explicar padrões de correlação, conduzirá a inferências inválidas sobre esses perfis.

No processo de construção e seleção de modelos, muitos pesquisadores e analistas preferem utilizar testes estatísticos para realizar comparações. Alguns deles são apresentados nas próximas subseções.

2.2.8.1 Teste da Razão de Verossimilhanças

O teste da razão de verossimilhanças (TRV) pode ser utilizado na comparação de modelos quando eles são ajustados pelos métodos de máxima verossimilhança (MR) ou máxima verossimilhança restrita (MVR). Esse teste baseia-se na comparação dos valores dos logaritmos das funções de verossimilhanças de modelos aninhados. Um modelo estatístico é denominado aninhado ou reduzido a um modelo de referência mais geral, quando ele pode ser obtido impondo-se determinadas restrições ao parâmetros do modelo de referência. No contexto de modelos lineares mistos, essas restrições podem ser impostas nas especificações de efeitos fixos, aleatórios e dos parâmetros de covariâncias (LITTELL et al., 2006; WEST; WELCH; GALECKI, 2007).

Segundo Pinheiro e Bates (2000), se L_2 é a função de verossimilhança do modelo de *referência* e L_1 é a função de verossimilhança do modelo *reduzido*, então temse que $L_2 > L_1$ e pela monotocidade da função logarítmica $ln(L_2) > ln(L_1)$. Assim, a estatística do TRV é dada por

$$2ln\left(\frac{L_2(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\theta}}_2; \boldsymbol{y})}{L_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\theta}}_1; \boldsymbol{y})}\right) = 2[ln(L_2(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\theta}}_2; \boldsymbol{y})) - ln(L_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\theta}}_1; \boldsymbol{y})], \quad (28)$$

que será positiva. Se k_i é o número de parâmetros a ser estimado no modelo i (i = 1, 2),

então assintoticamente ou para "grandes amostras" a distribuição da estatística do TRV, sob a hipótese nula de que o modelo reduzido é mais adequado, segue a distribuição χ^2 com $k_2 - k_1$ graus de liberdade. Além disso, para os casos em que os parâmetros encontram-se na fronteira do espaço paramétrico, a estatística do teste da razão de verossimilhanças segue uma mistura de χ^2 (SELF; LIANG, 1987; SHAPIRO, 1988; VERBEKE; MOLEN-BERGS, 2000).

Hipóteses sobre os parâmetros de um modelo linear misto podem ser testadas utilizando-se a estatística dada em (28). Se a hipótese nula for rejeitada, conclui-se que o modelo de *referência* (com mais parâmetros) é mais adequado. Caso contrário, as estimativas de L_1 e L_2 são próximas, indicando que o modelo *reduzido* com um menor número de parâmetros deve ser escolhido. Os seguintes testes da razão de verossimilhanças podem ser realizados:

• Teste para parâmetros de efeitos fixos

A importância dos termos fixos de um modelo linear misto pode ser verificada pelo teste da razão de verossimilhanças. Para tanto, os dois modelos, de referência e reduzido, devem diferir apenas em relação aos efeitos fixos e devem ser ajustados pelo método da máxima verossimilhança (MV). Dois modelos ajustados por máxima verossimilhança restrita (MVR) com matrizes de especificações de efeitos fixos X_i diferentes, mesmo que sejam aninhados, não podem ser comparados pelo TRV, pois na estimação por MVR é utilizada uma correção para as variâncias que depende da estrutura dos efeitos fixos e isso faz com que esses modelos não sejam comparáveis com relação à parte fixa (LITTELL et al., 2006; PINHEIRO; BATES, 2000; VERBEKE; MOLENBERGHS, 2000; WEST; WELCH; GALECKI, 2007).

A estatística do teste da razão de verossimilhanças para dois modelos aninhados com diferentes especificações de efeitos fixos é calculada como apresentado em (28) e segue, assintoticamente, a distribuição χ^2 com os graus de liberdade dado pela diferença do número de parâmetros de efeitos fixos entre os dois modelos.

• Teste para parâmetros de covariância

De acordo com West, Welch e Galecki (2007), o teste de razão de verossimilhanças para parâmetros de covariância assume que os modelos de *referência* e *reduzido* tenham o mesmo conjunto de efeitos fixos e diferentes conjuntos de parâmetros de covariância. Esses autores, além de outros como Littell et al. (2006) e Pinheiro e Bates (2000), recomendam que a estimação dos componentes de variância seja realizada pelo método da máxima verossimilhança restrita (MVR), dado ao fato desta estimação reduzir o viés inerente às estimativas obtidas por máxima verossimilhança.

Segundo Littell et al. (2006) e West, Welch e Galecki (2007), a realização do teste da razão de verossimilhanças, baseado no método da máxima verossimilhança restrita, deve ter uma atenção especial sob determinadas condições. Tais condições referem-se a parâmetros que estão na fronteira do espaço paramétrico, como por exemplo, quando se realiza um teste unilateral para um componente de variância sob as hipóteses $H_0: \sigma^2 = 0$ versus $H_a: \sigma^2 > 0$. Em casos como esse, a estatística do teste segue uma mistura de χ^2 , $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$.

2.2.8.2 Testes adicionais para parâmetros de efeitos fixos

De acordo com Verbeke e Molenberghs (2000) e conforme discutido na seção 2.2.4.1, o vetor de efeitos fixos é estimado por:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \boldsymbol{V}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \boldsymbol{V}_{i}^{-1} \boldsymbol{y}_{i},$$
(29)

em que o vetor desconhecido $\boldsymbol{\theta}$ de componentes de variância de \boldsymbol{D} e \boldsymbol{R}_i é substituído por suas estimativas MV ou MVR. Sob o modelo marginal (12) e condicionalmente a $\boldsymbol{\theta}$, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ segue uma distribuição normal multivariada com vetor de médias $\boldsymbol{\beta}$ e com matriz de variâncias e covariâncias dada por:

$$Var(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}^{\prime} \boldsymbol{V}_{i}^{-1} \boldsymbol{X}_{i}\right)^{-1}$$
(30)

em que, $V_i^{-1} = V_i^{-1}(\theta)$ e é estimada substituindo θ por suas estimativas MV ou MVR.

A fim de verificar a significância dos parâmetros de efeitos fixos do modelo, Verbeke e Molenbergs (2000) e West, Welch e Galecki (2007) discutem os testes adicionais t, Wald e Wald-F.

• Teste t

O teste t pode ser utilizado para testar cada parâmetro β_j , em β , j =

 $1, 2, \ldots, p$, sob as hipóteses:

$$H_0: \beta_j = 0$$
 versus $H_a: \beta_j \neq 0$,

em que a estatística t é dada por:

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

No contexto de modelos lineares mistos, a estatística t não segue uma distribuição t exata, sendo necessários métodos como os de Satterthwaite (1946) e Kenward e Roger (1997) para a determinação dos graus de liberdade associados. Sumariamente, esse teste verifica a necessidade de manter o parâmetro avaliado uma vez que os demais estão no modelo.

• Teste de Wald

O teste de Wald pode ser utilizado para testar hipóteses lineares da forma:

$$H_0: \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{0} \quad \text{versus} \quad H_a: \boldsymbol{L}\boldsymbol{\beta} \neq \boldsymbol{0},$$
 (31)

para uma matriz L conhecida. A estatística do teste é dada por:

$$W = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \boldsymbol{L}' \left[\boldsymbol{L} \left(\sum_{i=1}^{m} \boldsymbol{X}_{i}' \boldsymbol{V}_{i}^{-1} \left(\hat{\boldsymbol{\theta}} \right) \boldsymbol{X}_{i} \right)^{-1} \boldsymbol{L}' \right]^{-1} \boldsymbol{L} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right).$$

Sob H_0 , esta estatística segue, assintoticamente, uma distribuição χ^2 com graus de liberdade igual ao posto da matriz **L**.

Segundo Verbeke e Molenbergs (2000), a estatística do teste de Wald subestima a verdadeira variabilidade de $\hat{\boldsymbol{\beta}}$ por não considerar a variabilidade introduzida pela estimação dos componentes de variância ($\boldsymbol{\theta}$). Esse viés, frequentemente, é resolvido utilizando-se o teste t ou Wald-F.

• Teste Wald-F

O teste Wald-F tem suas hipóteses dadas como em (31) e a estatística dada por:

$$F = \frac{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \boldsymbol{L}' \left[\boldsymbol{L}\left(\sum_{i=1}^{m} \boldsymbol{X}_{i}' \boldsymbol{V}_{i}^{-1} \left(\hat{\boldsymbol{\theta}}\right) \boldsymbol{X}_{i}\right)^{-1} \boldsymbol{L}'\right]^{-1} \boldsymbol{L}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)}{posto(\boldsymbol{L})},$$

em que F segue uma distribuição F aproximada, com graus de liberdade do numerador dado pelo posto de L e do denominador, calculados por métodos de aproximação como definidos por Satterthwaite (1946) e Kenward e Roger (1997).

Os métodos de aproximação de graus de liberdade aplicados aos testes t e Wald-F levam em consideração a presença de efeitos aleatórios e resíduos correlacionados em um modelo linear misto. O método de Kenward–Roger, por exemplo, além do ajuste de graus de liberdade como realizado pelo método de Satterthwaite, também modifica a matriz de covariâncias estimadas para refletir as incertezas do uso de \hat{V}_i no lugar de V_i nas equações (29) e (30) (WEST; WELCH; GALECKI, 2007).

2.2.8.3 Critérios de informação

A seleção de modelos também pode ser realizada com auxílio de outras ferramentas como, por exemplo, os critérios de informação. Esses critérios levam em conta não apenas a qualidade do ajuste, mas também penalizam a inclusão de parâmetros extras. Assim, um modelo com mais parâmetros pode ter um ajuste melhor, mas não necessariamente será preferível em termos de critérios de informação, que consideram o princípio da parcimônia: "modelos mais simples devem ser escolhidos aos mais complexos, desde que a qualidade do ajuste seja similar".

Ao contrário do teste da razão de verossimilhanças, dois modelos quaisquer, aninhados ou não, podem ser comparados pelos critérios de informação. Os principais critérios são: Critério de Informação de Akaike (AIC) (AKAIKE, 1974), Critério de Informação de Akaike corrigido (AICc) (SUGIURA, 1978) e o Critério de Informação Bayesiano (BIC)(SCHWARZ, 1978). Quanto menores forem os valores de AIC, AICc ou BIC, melhor será o ajuste do modelo. Se os valores obtidos por esses critérios forem próximos para diferentes modelos ajustados, então a melhor opção será pelo modelo mais simples e parcimonioso.

• Critério de Informação de Akaike - AIC

O critério de Informação de Akaike (*Akaike's Information Criterion* - AIC) é baseado na teoria de decisão. Sua expressão é dada por:

$$AIC = -2 \log \left(L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}; \boldsymbol{y}) \right) + 2 p,$$

em que p representa o número total de parâmetros do modelo.

• Critério de Informação de Akaike corrigido - AICc

O desempenho do critério de Akaike pode ser insatisfatório se existirem muitos parâmetros em comparação ao tamanho da amostra (SUGIURA, 1978; SAKAMOTO; ISHIGURO; KITAGAWA, 1986). Sugiura (1978) derivou uma variante de segunda ordem do AIC, chamado de critério de informação de Akaike corrigido (AICc), dado por:

$$AICc = -2 \log \left(L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}; \boldsymbol{y}) \right) + 2p \left(\frac{n}{n-p-1} \right).$$

Segundo o mesmo autor, esta expressão pode ser reescrita equivalentemente

como:

$$AICc = AIC + \frac{2p(p+1)}{n-p-1},$$

em que n $(n = \sum_{i=1}^{m} n_i)$ é o tamanho amostral e p é o número total de parâmetros do modelo.

• Critério de Informação Bayesiano - BIC

O critério de informação bayesiano (*Bayesian Information Criterion* - BIC), também chamado de critério de informação de Schwarz (*Schwarz's Bayesian Information* *criterion* - SBC), é um critério semelhante ao AIC, porém penaliza mais os modelos com maior número de parâmetros. Sua expressão é dada por:

$$BIC = -2 \log \left(L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}; \boldsymbol{y}) \right) + p \ln(n),$$

em que p corresponde ao número de parâmetros e $n = \sum_{i=1}^{m} n_i$ o número de observações utilizadas para a estimação do modelo.

Assim como no teste da razão de verossimilhanças, os mesmos cuidados devem ser tomados ao utilizar os critérios de informação para comparar modelos ajustados pelo método de MVR. Nesse caso, dois modelos com efeitos fixos diferentes não são comparáveis, já que uma mudança na matriz \boldsymbol{X}_i resulta em uma mudança em $L_{MRV}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}; \boldsymbol{y})$. A comparação entre modelos com diferentes especificações de efeitos fixos pode ser realizada pelos critérios de informação desde que tenha sido utilizado o método de MV durante o processo de estimação dos modelos (LITTELL et al. 2006; PINHEIRO; BATES, 2000).

2.2.8.4 Estratégias para construção de modelos

Como mencionado no início da subseção 2.2.8, os processos de construção e seleção de modelos nem sempre são fáceis de serem executados, principalmente no caso de modelos lineares mistos. Além das especificações das partes fixa e aleatória, existem diversas formas de especificar as estruturas de covariâncias para D e R_i , que são responsáveis pela variação da parte aleatória. Portanto, vários passos no ajuste, investigação e seleção são necessários na construção de um modelo linear misto, de forma que o modelo final seja capaz de representar bem a média e a variabilidade presente nos dados observados.

Com relação à identificação da estrutura de covariâncias intraunidades amostrais (\mathbf{R}_i) em estudos longitudinais, Singer, Nobre e Rocha (2012) sugerem vários procedimentos dentre os quais:

- Análise visual do gráfico de perfis médios para propor um modelo para os efeitos fixos;
- 2. Análise visual dos gráficos de perfis individuais e de perfis individuais centralizados

para identificar padrões que indiquem possíveis efeitos aleatórios e heterocedasticidade ao longo das condições de avaliação;

- Análise das matrizes de covariâncias e correlações intraunidades amostrais, quando for possível calculá-las, para identificar seus padrões de variação;
- 4. Utilização dos critérios de informação como ferramenta auxiliar para a seleção das estruturas de covariância intraunidades amostrais;
- 5. Análise de resíduos

Segundo West, Welch e Galecki (2007), a construção de modelos, usualmente, envolve um equilíbrio entre conceitos estatísticos e considerações sobre o fenômeno estudado. Várias estratégias podem ser utilizadas para cada aplicação na busca do melhor modelo. Duas delas são destacadas pelos autores.

• Estratégia top-down

Sugerida por Verbeke e Molenbergs (2000), essa estratégia parte do pressuposto de que a estrutura de covariâncias V_i é responsável por modelar toda variabilidade presente nos dados que não poder ser explicada pelos efeitos fixos. Desta forma, primeiramente ajusta-se um modelo maximal com todos os termos de efeitos fixos possíveis e de suas interações, retirando toda forma de variação sistemática presente nos dados observados.

Em um segundo passo, seleciona-se um conjunto dos possíveis efeitos aleatórios que podem ser incluídos no modelo. Como visto no teste para parâmetros de covariância descrito na subseção 2.2.8.1, a decisão entre incluir ou não um efeito aleatório selecionado pode ser realizada utilizando-se o teste da razão de verossimilhanças pelo método MVR, tomando os devidos cuidados para parâmetros na fronteira do espaço paramétrico e respeitando a hierarquia no caso de modelos polinomiais, testando-se, primeiramente, o efeito aleatório de maior ordem e mantendo termos de ordens inferiores, caso significativo. Nesse passo, diferentes classes de matrizes positivas definidas podem ser comparadas a fim de se encontrar a melhor estrutura da matriz de covariâncias dos efeitos aleatórios D. Após inserir os efeitos fixos (β) e aleatórios (b_i) no modelo linear misto, a variabilidade restante dos dados observados deve-se exclusivamente aos erros ε_i . O próximo passo consiste em especificar, condicionalmente aos efeitos aleatórios, uma estrutura apropriada para a matriz de erros intra-unidades experimentais, R_i . Diferentes estruturas são possíveis nesse passo e a mais apropriada pode ser selecionada pelo teste da razão de verossimilhanças com os parâmetros estimados pelo MVR ou com auxílio dos critérios de informação.

De acordo com todas as especificações anteriores, uma possível redução do modelo deve ser investigada. Testes sobre os efeitos aleatórios incluídos no segundo passo podem ser realizados a fim de identificar se realmente são necessários no modelo. Além disso, dado que a estrutura de covariâncias final V_i já foi selecionada, uma possível redução nos efeitos fixos deve ser investigada com auxílio de testes apropriados como os discutidos nas subseções 2.2.8.1 e 2.2.8.2.

• Estratégia step-up

Ao contrário da *top-down*, esta estratégia descrita por Snijders e Bosker(1999) e por Raudenbush e Bryk (2002) inicia-se com o ajuste de um modelo simples, contendo apenas o efeito da média geral e de algum possível efeito aleatório. Em seguida, são incluídos efeitos de covariáveis fixas e de possíveis efeitos aleatórios associados. Por fim, escolhe-se a matriz de covariâncias mais apropriada.

A decisão entre incluir ou não termos no modelo pode ser realizada com auxílio do teste da razão de verossimilhanças e dos critérios de informação (PINHEIRO; BATES, 2000).

2.2.9 Diagnósticos

Conforme apresentado, o modelo linear misto (1) descrito no início da subseção 2.2 é construído sobre determinadas suposições. Antes de fazer inferências sobre o modelo ajustado, é necessário verificar se essas suposições são satisfeitas, a fim de evitar falhas sistemáticas e consequentemente conclusões duvidosas.

Técnicas formais ou informais podem ser usadas para a verificação do ajuste de um modelo a um conjunto de dados. As formais envolvem testes de hipóteses e medidas de qualidade do ajuste, dentre outras. As informais dependem da habilidade adquirida e baseiam-se em análises gráficas visuais para a detecção de padrões específicos ou de pontos discrepantes.

Em um modelo linear clássico, as técnicas de diagnósticos estão bem estabelecidas na literatura estatística (BARBIN, 2013; PIMENTEL GOMES, 2009; SHAPIRO; WILK, 1965; TUKEY, 1949). Por exemplo, a análise gráfica de resíduos é usada para verificar pressuposições tal como homogeneidade de variância e independência, além de indicar a necessidade de uma possível transformação dos dados. A qualidade do ajuste é verificada por meio de medidas resumos e estatísticas calculadas com base nos resíduos. O inter-relacionamento entre efeitos fixos e seu impacto na análise é estudado com diagnósticos de colinearidade, ao passo que a importância e peso de observações individuais na análise são verificadas por meio de medidas de influência (LITTELL et al., 2006).

Por outro lado, a presença de efeitos aleatórios e de diferentes estruturas de covariâncias em modelos lineares mistos aumentam a complexidade e dificultam a realização e interpretação de métodos de diagnósticos. Além disso, esses métodos devem ser aplicados ao longo de todo o processo de construção dos modelos mistos, não se restringindo apenas ao modelo do ajuste final (WEST; WELCH; GALECKI, 2007).

Segundo Pinheiro e Bates (2000), dentre os métodos utilizados para avaliar as pressuposições do modelo linear misto, os mais úteis são aqueles baseados em análises gráficas de resíduos, de valores estimados e de efeitos aleatórios preditos. De acordo com os autores, testes de hipóteses podem ser aplicados formalmente para verificar a validade das pressuposições, porém são raros os casos em que as conclusões obtidas por esses testes contradizem as informações apresentadas em diagnósticos gráficos.

2.2.9.1 Resíduos

Os resíduos têm papel fundamental na verificação do ajuste de um modelo. Eles são utilizados para verificar as pressuposições, visualizar padrões de comportamento ou detectar pontos discrepantes (*outliers*) e observações potencialmente influentes.

Em modelos lineares mistos, como visto em (3) na subseção 2.2, o vetor de respostas \boldsymbol{Y} tem distribuição marginal normal multivariada com vetor de médias $\boldsymbol{X\beta}$ e matriz de covariâncias $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$ e o vetor de respostas considerado condicionalmente aos efeitos aleatórios $\boldsymbol{Y}|\boldsymbol{b}$ tem distribuição condicional com média $\boldsymbol{X\beta} + \boldsymbol{Zb}$ e variância \boldsymbol{R} . Isso distingue dois tipos de resíduos:

- (i) Resíduos marginais: $\hat{\boldsymbol{\varepsilon}}^* = \boldsymbol{Y} \boldsymbol{X}\hat{\boldsymbol{\beta}}$, que predizem os erros marginais $\boldsymbol{\varepsilon}^* = \boldsymbol{Y} E(\boldsymbol{Y})$;
- (ii) Resíduos condicionais: $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} \boldsymbol{X}\hat{\boldsymbol{\beta}} \boldsymbol{Z}\hat{\boldsymbol{b}}$, que predizem os erros condicionais $\boldsymbol{\varepsilon} = \boldsymbol{Y} - E(\boldsymbol{Y}|\boldsymbol{b}).$

Os resíduos marginais ou condicionais não são adequados para verificar pressuposições relacionadas ao modelo. Segundo West, Welch e Galecki (2007), mesmo que os resíduos do modelo sejam não correlacionados e tenham variâncias homogêneas, os resíduos condicionais tenderão a ser correlacionados e suas variâncias poderão ser diferentes para subgrupos diferentes de unidades amostrais.

Problemas com a interpretação de resíduos com variâncias desiguais podem ser evitados aplicando-se diferentes padronizações. Ao dividir os resíduos pela estimativa do desvio padrão da variável dependente obtém-se os resíduos de Pearson, que são apropriados apenas se a variabilidade de $\hat{\beta}$ puder ser ignorada. Teoricamente, resíduos padronizados são obtidos pela divisão dos resíduos pelos seus respectivos erros padrões. Porém, na prática, tais desvios são desconhecidos e então a divisão é feita pelos seus respectivos desvios padrões estimados, obtendo-se assim os resíduos estudentizados (WEST; WELCH; GALECKI, 2007).

O resíduo marginal estudentizado é dado por:

$$\hat{\boldsymbol{\varepsilon}}^*{}_e = \frac{\hat{\boldsymbol{\varepsilon}}^*}{\sqrt{\hat{Var}(\hat{\boldsymbol{\varepsilon}}^*)}}.$$
(32)

E o resíduo condicional estudentizado por:

$$\hat{\boldsymbol{\varepsilon}}_e = \frac{\hat{\boldsymbol{\varepsilon}}}{\sqrt{\hat{Var}(\hat{\boldsymbol{\varepsilon}})}}.$$
(33)

Segundo Littell et al. (2006), resíduos marginais ou condicionais têm diferentes padronizações dado que as variâncias de $\hat{\epsilon}^*$ e $\hat{\epsilon}$ são diferentes. Além disso, dois tipos de cálculo são possíveis para resíduos estudentizados. Se a observação referente ao resíduo em questão é incluída no cálculo, tem-se o resíduo estudentizado internamente, caso contrário, tem-se o resíduo estudentizado externamente. Outro ponto destacado pelos autores é a grande diferença entre as propriedades do resíduo marginal e condicional e como o resíduo marginal em modelos lineares mistos decorrente da matriz de variâncias e covariâncias V = ZGZ' + R pode se comportar diferentemente do resíduo de um modelo linear clássico.

Em seus trabalhos, Pinheiro e Bates (2000) e West, Welch e Galecki (2007) utilizaram um conjunto rico e integrado de gráficos de diagnósticos para verificar as pressuposições do modelo. Padrões de correlação e heterocedasticidade nos resíduos são diagnosticados por meio do gráfico de resíduos condicionais padronizados versus valores estimados. A pressuposição de normalidade para os resíduos condicionais padronizados é testada por meio de gráficos de quantis normais (Q-Q plot) e a qualidade do ajuste final de cada modelo é verificada por meio do gráfico de valores estimados versus observados.

2.3 A cultura da alfafa

A alfafa (*Medicago sativa* L.), originária da Ásia e do sul do Cáucaso, é uma leguminosa forrageira capaz de se adaptar a diferentes condições de clima e solo, o que ajudou na sua disseminação por diversas regiões agrícolas do mundo (MOREIRA et al., 2007). Além da sua boa adaptabilidade, a alfafa reúne características importantes relacionadas à produtividade, palatabilidade, digestibilidade, nutrição e qualidade proteica (RASSINI; FERREIRA; CAMARGO, 2008). Segundo Moreira et al. (2007), o interesse pelo cultivo da alfafa está relacionado à sua qualidade como planta forrageira, possuindo altos níveis de proteínas, cálcio, fósforo e das vitaminas A e E. Por todas essas qualidades, a alfafa é conhecida como "rainha das forrageiras" (NUERNBERG, 1986; HIJANO; NAVARRO, 1995; RASSINI, 1998).

No Brasil, os primeiros registros do cultivo da alfafa datam do ano de 1850, quando foi introduzida no Rio Grande do Sul por imigrantes vindos da Argentina e do Uruguai (SAIBRO, 1985). É utilizada, principalmente, na alimentação de cavalos de raça, onde o custo com a alimentação não é tão relevante. Em rebanhos leiteiros a relação custo benefício é favorável apenas quando se tem animais de alto valor zootécnico. Segundo Rassini, Ferreira e Camargo (2008), é uma das culturas mais importantes para a alimentação de rebanhos leiteiros especializados e pode ser oferecida aos animais sob a forma conservada (feno e silagem) ou na forma *in natura* (verde) picada ou, ainda, ser utilizada em pastejo rotativo, conseguindo-se excelentes resultados em termos de produção leiteira. Apesar do seu bom desempenho e de sua excelente qualidade nutritiva quando comparada a outras plantas forrageiras, a expansão para as regiões Sudeste e Centro Oeste do país tem sido restrita devido, principalmente, à sazonalidade da pecuária leiteira e à pressão de culturas agrícolas por novas áreas (cana-de-açúcar, citros, eucalipto, dentre outras). Segundo Rassini, Ferreira e Camargo (2008), entre outros fatores que dificultam a expansão da cultura no Brasil, também se destacam a limitada produção de sementes, a pequena disponibilidade de cultivares adaptadas aos trópicos e a falta de conhecimento das exigências da cultura por parte dos produtores.

Tal cultura exige grandes investimentos para atingir produtividade e qualidade nutricional satisfatórias. Os principais custos estão relacionados à formação da área de cultivo, envolvendo processos como preparo e correção da fertilidade do solo, aquisição de sementes certificadas e plantio, estendendo-se ao longo da manutenção da forragem com o controle de plantas daninhas, doenças e pragas, adubação, irrigação adequada etc. (RASSINI; FERREIRA; CAMARGO, 2008).

A produtividade e qualidade da alfafa estão diretamente relacionadas à fertilidade do solo e ao fornecimento de nutrientes essenciais à cultura. Segundo Moreira, Bernardi e Rassini (2008), a correção da acidez do solo aliado ao fornecimento adequado de nutrientes garantem cerca de 50% de aumento da produtividade. A utilização de fertilizantes e corretivos deve ser realizada a partir dos resultados obtidos da análise de solo. O rendimento dessa forragem é significativamente influenciado pelo uso de calcário e fertilizantes contendo macronutrientes como cálcio (Ca), magnésio (Mg), enxofre (S), nitrogênio (N), potássio (K) e fósforo (P). Em quantidades menores, mas não menos importantes, os micronutrientes boro (B), cobre (Cu), ferro (Fe), manganês (Mn), molibdênio (Mo), niquel (Ni) e zinco (Zn) também são essenciais para as plantas.

Dentre esses micronutrientes, a alfafa é responsiva ao teor de boro no solo e a deficiência desse micronutriente pode reduzir a produtividade das lavouras, ainda mais em solos brasileiros cujo teor de boro geralmente é baixo (RAIJ, 1991). Contudo, a utilização de boro demanda cuidados especiais, uma vez que a sua aplicação em excesso pode ser tóxica às plantas. Isso justifica a necessidade de se avaliar os níveis deste elemento no solo e na planta, e assim determinar as quantidades corretas a serem aplicadas (FAGERIA, 2000).

3 MATERIAL E MÉTODOS

3.1 Material

No presente estudo, foram analisados os dados de produção de matéria seca de 6 cortes sequenciais de avaliação da cultura da alfafa (*Medicago sativa* L., cultivar Crioula). Dois experimentos foram realizados em 2008 no campo experimental da Embrapa Pecuária Sudeste, em São Carlos, Estado de São Paulo (22° 01' 10" S, 47° 53' 38" O), em Latossolo Vermelho Amarelo distrófico, sendo um experimento realizado sob saturação por base de 60% e o outro sob saturação por base de 80% (Figura 1).



Figura 1 – Ensaio com a cultura da alfafa realizado no campo experimental da Embrapa Pecuária Sudeste

Visando definir os níveis críticos de boro (B) na cultura da alfafa (*Medicago sativa* L., cultivar Crioula) cultivada nas condições edafoclimáticas locais foram utilizadas duas fontes de boro (ácido bórico (H₃BO₃) com 17% de B e Ulexita (Na₂.2CaO.5B₂O₃.16H₂O) com 9 a 10% de B) associadas à quatro doses (1, 3, 6 e 9 kg ha⁻¹). Além disso, um tratamento controle, sem boro, também foi utilizado a fim de realizar comparações, ou seja, servir como referência (padrão) para avaliação dos demais tratamentos e para obtenção de informações complementares.

Nos dois experimentos, os oito tratamentos compostos pelas combinações entre as duas fontes e as quatro doses de boro (esquema fatorial 2×4) e o tratamento controle foram aleatorizados a quatro blocos. Em cada unidade experimental, composta por uma parcela de 8 m² de alfafa, foram realizados seis cortes com intervalo de 30 dias (fator longitudinal), sendo que o primeiro corte ocorreu três meses após o plantio. Diversas variáveis foram obtidas a partir do material vegetal coletado em cada corte, porém apenas a variável produção de matéria seca (kg ha⁻¹) na saturação por base 60% foi analisada neste trabalho.

Exceto o nitrogênio (N) e o boro (B), as adubações foram feitas de acordo com Moreira et al. (2006). Os micronutrientes cobre (Cu), ferro (Fe), manganês (Mn) e zinco (Zn) e as doses de boro correspondentes aos tratamentos foram misturados e aplicados juntamente com o superfosfato simples (20% de P_2O_5) e incorporados com enxada rotativa.

3.2 Métodos

3.2.1 Identificação dos tratamentos

Primeiramente, para que os dados fossem analisados adequadamente, foram identificados os fatores presentes e a composição dos tratamentos no experimento de adubação com boro em alfafa. Os tratamentos foram obtidos por meio da composição que pode ser observada na Tabela 1.

Tabela 1 - Composição dos tratamentos no experimento de adubação com boro em alfafa

Fonte \setminus Dose (kg ha ⁻¹)	Controle	1	3	6	9
Controle	T_1				
Ácido bórico		T_2	T_3	T_4	T_5
Ulexita		T_6	T_7	T_8	T_9

Na composição dos tratamentos existem dois conjuntos disjuntos: o conjunto que contém apenas o tratamento controle (T_1) e o conjunto que contém os tratamentos formados pelo esquema fatorial $(T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9)$. Deste modo, a fim de testar o tratamento controle *versus* tratamentos do esquema fatorial, criou-se um novo fator, chamado *Adicional*, com dois níveis (*Adicional* = 0 se controle, e 1 se fatorial), em que os tratamentos ficaram aninhados. Por definição, dois fatores são considerados aninhados se unidades experimentais com o mesmo nível de um fator aninhado, mas diferentes níveis do fator que o aninha, não possuem características em comum (BRIEN, 2015).

A combinação dos nove tratamentos com os quatro blocos formam as 36

unidades experimentais presentes no experimento.

3.2.2 Análise exploratória

A fim de visualizar padrões de comportamento, foram construídos gráficos dos perfis individuais por unidade experimental da produção de matéria seca ao longo dos dias de avaliação e dos perfis médios para cada um dos tratamentos presentes no estudo. Adicionalmente, foram construídos gráficos dos perfis médios (com barras de erros padrões) distinguindo as fontes e doses com intuito de visualizar o comportamento médio dentro dos fatores e seus níveis. Todos os gráficos dos perfis individuais e médios apresentados na análise exploratória dos dados foram obtidos por implementações utilizando as bibliotecas lattice versão 0.20-31 (SARKAR, 2015) e ggplot2 versão 1.0.1 (WICKHAM; CHANG, 2015) disponíveis para o programa R (R *Development Core Team*, 2015) e o procedimento SGPLOT disponível no programa SAS versão 9.3 (SAS INSTITUTE INC., 2015).

3.2.3 Diagrama de Hasse e estratégia *top-down* na especificação do modelo estatístico

Devido à complexidade do delineamento, que envolve dados longitudinais em ensaio fatorial com tratamento adicional, optou-se pela utilização do diagrama de Hasse para auxiliar na compreensão e visualização simbólica do relacionamento de todos os fatores presentes no estudo, obtenção dos números de graus de liberdade associados a cada fonte de variação e construção do modelo. De acordo com os procedimentos propostos por Brien (2015), separaram-se fatores não aleatorizados dos aleatorizados descrevendo suas estruturas experimentais e suas respectivas fontes derivadas.

O diagrama de Hasse foi construído de acordo com as regras descritas detalhadamente por Alcarde (2007, 2012) e Brien (2015) a partir das estruturas experimentais e de suas fontes derivadas. O fator universal, representado pela média geral, foi incluído no topo do diagrama. Seguindo uma hierarquia, nas linhas inferiores foram colocados os outros fatores de modo que, dado a posição de um determinado fator, todos os que ficaram acima são marginais a ele. A conexão foi realizada desenhando uma seta para cima. Fatores cruzados ficaram no mesmo nível horizontalmente. Do lado direito de cada fator generalizado adicionou-se sua fonte. Do lado esquerdo do fator escreveu-se o número de níveis e do lado direito da fonte o número de graus de liberdade provenientes da diferença entre o número de níveis e a soma dos graus de liberdade de todos os fatores generalizados marginais a esse fator.

A estratégia top-down (VERBEKE; MOLENBERGHS, 2000) descrita na subseção 2.2.8.4 também foi adotada para auxiliar nas especificações dos modelos estatísticos neste estudo. Inicialmente, considerou-se uma abordagem qualitativa para todos os fatores de um modelo linear misto maximal com todos os termos e interações de efeitos fixos possíveis. O modelo linear misto (34) correspondente à produção de matéria seca $y_{fd(a)cj}$ avaliado no f-ésimo nível de fonte e d-ésimo nível de dose aninhados no a-ésimo nível do fator adicional criado, no c-ésimo nível de corte e no j-ésimo nível de bloco é expresso para $f=1,2, d=1,2,\ldots,4, a=1,2, c=1,2,\ldots,6$ e $j=1,2,\ldots,4$ como:

$$y_{fd(a)cj} = \underbrace{\mu + \alpha_a + \phi_{f(a)} + \delta_{d(a)} + (\phi\delta)_{fd(a)} + \chi_c + (\alpha\chi)_{ac} + (\phi\chi)_{f(a)c} + (\delta\chi)_{d(a)c} + (\phi\delta\chi)_{fd(a)c}}_{\text{efeitos aleatórios}},$$
(34)

em que μ é uma constante comum a todas as observações, α_a é o efeito fixo do fator adicional criado para comparar tratamento controle *versus* tratamentos do esquema fatorial, $\phi_{f(a)}$ é o efeito fixo da fonte aninhado ao fator adicional; $\delta_{d(a)}$ é o efeito fixo da dose aninhado ao fator adicional, $(\phi\delta)_{fd(a)}$ é o efeito fixo da interação entre fonte e dose aninhado ao fator adicional, χ_c é o efeito fixo de corte, $(\alpha\chi)_{ac}$ é o efeito fixo da interação entre o fator adicional e corte, $(\phi\chi)_{f(a)c}$ é o efeito fixo da interação entre fonte aninhado ao fator adicional e corte, $(\phi\chi)_{d(a)c}$ é o efeito fixo da interação entre dose aninhado ao fator adicional e corte, $(\delta\chi)_{d(a)c}$ é o efeito fixo da interação entre dose aninhado ao fator adicional e corte, $(\phi\delta\chi)_{fd(a)c}$ é o efeito fixo da interação entre dose aninhado ao fator adicional e corte, $(\phi\delta\chi)_{fd(a)c}$ é o efeito fixo da interação entre fonte interagindo com dose aninhado ao fator adicional e corte; b_j é o efeito aleatório de bloco; $b_{fd(a)j}$ é o efeito aleatório de unidade experimental e $\varepsilon_{fd(a)cj}$ é o erro associado à observação $y_{fd(a)cj}$.

3.2.4 Seleção dos efeitos aleatórios

Com relação à parte de efeitos aleatórios, assume-se que o vetor de efeitos aleatórios, \boldsymbol{b}_i , associado à *i*-ésima unidade experimental (i = 1, 2, ..., 36) segue uma distribuição normal bivariada:

$$oldsymbol{b_i} = \left[egin{array}{c} b_j \ b_{fd(a)j} \end{array}
ight] , \qquad oldsymbol{b_i} \sim N_2\left(oldsymbol{0},oldsymbol{D}
ight).$$

Cada um dos dois efeitos aleatórios tem média zero e a matriz de variâncias e covariâncias D é dada por:

$$\boldsymbol{D} = Var(\boldsymbol{b}_i) = \begin{bmatrix} \sigma_{b_j}^2 & 0\\ 0 & \sigma_{b_{fd(a)j}}^2 \end{bmatrix}$$

De acordo com o segundo passo da estratégia *top-down*, a necessidade de manter ou não um efeito aleatório foi realizada com base no teste da razão de verossimilhanças (TRV) com os parâmetros estimados pelo método da máxima verossimilhaça restrita (MVR), tomando-se os devidos cuidados para parâmetros na fronteira do espaço paramétrico, ou seja, nestes casos foi calculada a razão de verossimilhanças entre o modelo de referência e o modelo aninhado, testando-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 de efeitos aleatórios (b_i) do modelo aninhado. Além do TRV, os critérios de informação AIC, AICc e BIC também foram usados b na tomada de decisão.

À medida que foram retirados os efeitos aleatórios do modelo (34), novos modelos foram criados. O modelo aninhado (35) sem o efeito aleatório associado às unidades experimentais é dado por:

$$y_{fd(a)cj} = \overbrace{\mu + \alpha_a + \phi_{f(a)} + \delta_{d(a)} + (\phi\delta)_{fd(a)} + \chi_c + (\alpha\chi)_{ac} + (\phi\chi)_{f(a)c} + (\delta\chi)_{d(a)c} + (\phi\delta\chi)_{fd(a)c}}^{\text{efeitos fixos}} + \underbrace{b_j + \varepsilon_{fd(a)cj}}_{\text{efeitos aleatórios}}.$$
(35)

A hipótese nula pode ser escrita como:

Hipótese 3.2.1. H_0 : O efeito aleatório associado com as unidades experimentais pode ser omitido do modelo 34.

Simbolicamente, as hipóteses nula e alternativa podem ser escritas da seguinte forma:

$$\boldsymbol{H_0}: \boldsymbol{D} = \begin{bmatrix} \sigma_{b_j}^2 & 0\\ 0 & 0 \end{bmatrix} \quad vs \quad \boldsymbol{H_1}: \boldsymbol{D} = \begin{bmatrix} \sigma_{b_j}^2 & 0\\ 0 & \sigma_{b_{fd(a)j}}^2 \end{bmatrix}$$

Também foi proposto o modelo aninhado (36) sem efeitos aleatórios, exceto

pelo erro $\varepsilon_{fd(a)cj}$, apenas com o intuito de destacar a importância da inclusão de efeitos aleatórios no modelo:

$$y_{fd(a)cj} = \underbrace{\mu + \alpha_a + \phi_{f(a)} + \delta_{d(a)} + (\phi\delta)_{fd(a)} + \chi_c + (\alpha\chi)_{ac} + (\phi\chi)_{f(a)c} + (\delta\chi)_{d(a)c} + (\phi\delta\chi)_{fd(a)c}}_{\text{efeito aleatório}},$$
(36)

A hipótese nula pode ser escrita como:

Hipótese 3.2.2. H_0 : Os efeitos aleatórios associados ao bloco e a unidade experimental podem ser omitidos do modelo 34.

Simbolicamente, as hipóteses nula e alternativa podem ser escritas da seguinte forma:

$$\boldsymbol{H_0}: \boldsymbol{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad vs \quad \boldsymbol{H_1}: \boldsymbol{D} = \begin{bmatrix} \sigma_{b_j}^2 & 0 \\ 0 & \sigma_{b_{fd(a)j}}^2 \end{bmatrix}$$

Os modelos (34), (35) e (36) foram ajustados pelo método da máxima verossimilhança restrita (MVR), conforme descrito na subseção 2.2.4.2. As estimativas MVR foram obtidas por meio do algoritmo *Ridge-stabilized Newton-Raphson*. Uma breve discussão sobre as vantagens da utilização desse algoritmo podem ser encontradas em Littell et al. (2006). Todos os ajustes foram realizados por meio do procedimento PROC MIXED disponível no programa SAS versão 9.3 (SAS INSTITUTE INC., 2015) com auxílio da macro %RMMS (WANG, 2006) do SAS.

3.2.5 Seleção da estrutura de covariâncias para os resíduos

Selecionado o modelo com estrutura mais adequada em relação à parte aleatória, seguindo a estratégia *top-down*, testou-se a incorporação de diferentes estruturas (descritas na subseção 2.2.1) para a matriz de erros intra-unidades experimentais (\mathbf{R}_i). O modelo com a estrutura mais apropriada foi selecionado pelo teste da razão de verossimilhanças (TRV) com os parâmetros estimados pelo método da máxima verossimilhaça restrita (MVR). No processo de seleção de modelos, além da significância do TRV, também foi considerado o modelo mais parcimonioso e os critérios de informação AIC, AICc e BIC, descritos na subseção 2.2.8.3, que levam em conta não apenas a qualidade do ajuste, mas também penalizam a inclusão de parâmetros extras.

Todos os ajustes foram realizados por meio do procedimento PROC MIXED disponível no programa SAS versão 9.3 (SAS INSTITUTE INC., 2015) com auxílio da macro %RMMS (WANG, 2006) do SAS.

3.2.6 Teste para efeitos fixos

Após a seleção de efeitos aleatórios da matriz D e a identificação da estrutura adequada para a matriz \mathbf{R}_i e, consequente definição da matriz \mathbf{V}_i , realizou-se o teste para efeitos fixos por meio do teste Wald-F descrito na subseção 2.2.8.2, onde se testou a importância de cada um dos termos fixos e suas interações.

3.2.7 Estudo de regressão

Assim como nos trabalhos de Barbin (2013) e Pimentel Gomes (2009), como continuidade da abordagem qualitativa, realizou-se também uma abordagem quantitativa (estudo de regressão) para o fator longitudinal corte. Cada um dos graus de liberdade associados ao fator em questão foi isolado, a fim de avaliar separadamente os efeitos de primeiro grau ou linear, de segundo grau ou quadrático, de terceiro grau ou cúbico, de quarto grau e de quinto grau. De acordo com essa abordagem quantitativa, o modelo linear misto (37) correspondente para produção de matéria seca $y_{fd(a)j}$ avaliada no fésimo nível de fonte e d-ésimo nível de dose aninhados no a-ésimo nível do fator adicional criado e, no j-ésimo nível de bloco é expresso para $f=1,2, d=1,2,\ldots,4, a=1,2$ e $j=1,2,\ldots,4$ como:

$$y_{fd(a)j} = \underbrace{\mu + \alpha_a + \phi_{f(a)} + \delta_{d(a)} + (\phi\delta)_{fd(a)} + \beta_{fd(a)1}C + \beta_{fd(a)2}C^2 + \beta_{fd(a)3}C^3 + \beta_{fd(a)4}C^4 + \beta_{fd(a)5}C^5}_{\text{efeitos aleatórios}} + \underbrace{b_j + b_{fd(a)j} + \varepsilon_{fd(a)j}}_{\text{efeitos aleatórios}},$$
(37)

em que μ é uma constante comum a todas as observações, α_a é o efeito fixo do fator adicional criado para comparar tratamento controle *versus* tratamentos do esquema fatorial, $\phi_{f(a)}$ é o efeito fixo da fonte aninhado no fator adicional; $\delta_{d(a)}$ é o efeito fixo da dose aninhado no fator adicional, $(\phi\delta)_{fd(a)}$ é o efeito fixo da interação entre fonte e dose aninhado no fator adicional, $\beta_{fd(a)1}$ coeficiente associado ao efeito linear de corte (C), $\beta_{fd(a)2}$ é coeficiente associado ao efeito quadrático de corte (C^2) , $\beta_{fd(a)3}$ coeficiente associado ao efeito cúbico de corte (C^3) , $\beta_{fd(a)4}$ é coeficiente associado ao efeito de quarto grau de corte (C^4) , $\beta_{fd(a)5}$ é coeficiente associado ao efeito de quinto grau de corte (C^5) ; b_j é o efeito aleatório de bloco; $b_{fd(a)j}$ é o efeito aleatório de unidade experimental e $\varepsilon_{fd(a)j}$ é o erro associado à observação $y_{fd(a)j}$.

Após a escolha do modelo envolvendo o estudo de regressão, assim como realizado na abordagem qualitativa, testou-se a incorporação de diferentes estruturas para a matriz de erros intra-unidades experimentais (\mathbf{R}_i). Para tanto, considerou-se o resultado do teste da razão de verossimilhanças, o modelo mais parcimonioso e os critérios de informação AIC, AICc e BIC.

Após a identificação da estrutura adequada para a matriz \mathbf{R}_i realizou-se o teste para efeitos fixos no estudo de regressão por meio do teste Wald-F, onde foi verificada a importância de cada um dos componentes de regressão.

3.2.7.1 Diagnósticos

Assim como nos trabalhos de Littell et al. (2006), Pinheiro e Bates (2000) e West, Welch e Galecki (2007), utilizou-se um conjunto de gráficos de diagnósticos para verificar as pressuposições do modelo final. Padrões de correlação e heterocedasticidade nos resíduos foram investigados por meio de análises gráficas dos resíduos condicionais estudentizados versus valores ajustados. Padrões de periodicidade dos resíduos condicionais estudentizados ao longo do fator longitudinal também foram investigados. A pressuposição de normalidade para os resíduos condicionais estudentizados foi confirmada por meio do gráfico de quantis da normal padronizada (Q-Q plot) e a qualidade do ajuste final do modelo foi verificada por meio do gráfico de valores ajustados versus observados.

Os gráficos apresentados na análise de diagnóstico para avaliar a adequacidade do modelo ajustado foram obtidos por implementações utilizando as procedimentos GPLOT e SGPLOT disponíveis no programa SAS versão 9.3 (SAS INSTITUTE INC., 2015).

Anexo a este trabalho, encontram-se todos os códigos computacionais implementados nos programas **R** versão 3.2.1 (R *Development Core Team*, 2015) e SAS versão 9.3 (SAS INSTITUTE INC., 2015) utilizados para desenvolver a metodologia proposta.

4 RESULTADOS E DISCUSSÃO

4.1 Análise exploratória

Os perfis individuais por unidade experimental da produção de matéria seca ao longo dos dias de avaliação são apresentados na Figura 2. Cada unidade experimental apresentou um comportamento específico, sendo notória a variabilidade presente na produtividade de uma unidade para outra com relação ao intercepto, o que sugeriu a inclusão de efeito aleatório a esse parâmetro. Em seus estudos de avaliação do conforto animal em aviários, Sercundes (2014) também utiliza o gráfico de perfis individuais de cada unidade experimental para sugerir a inclusão de efeitos aleatórios aos parâmetros associados ao intercepto, coeficiente linear e quadrático de um modelo polinomial capaz de representar as possíveis variações de cada unidade experimental ao longo de observações realizadas no tempo. Além disso, Kéry (2010) simulou conjuntos de dados com coeficientes aleatórios associados ao intercepto e coeficiente linear de um modelo linear misto a fim de ilustrar comportamentos específicos de populações de serpentes (*Vipera aspis*).



Figura 2 – Perfis individuais da produção de matéria seca da parte aérea (kg ha⁻¹) de plantas de alfafa, por unidade experimental ao longo dos dias de avaliação

Os perfis médios da produção de matéria seca da parte aérea de plantas de alfafa para cada um dos tratamentos ao longo dos dias de avaliação são apresentados
na Figura 3. Observa-se que o período de maior média apresenta grande diferença para determinados tratamentos, como por exemplo para o tratamento com ácido bórico e dose 4 kg ha⁻¹ (T₄), que apresenta maior média aos 210 dias, período superior a todos os outros tratamentos. Observa-se também que a oscilação média de produtividade ao longo do tempo pode estar associado a um polinômio de grau superior, capaz de explicar o comportamento médio com maior precisão.



Figura 3 – Perfis médios da produção de matéria seca da parte a
érea (kg $\rm ha^{-1})$ de plantas de alfafa, por tratamento a
o longo dos dias de avaliação

Os perfis médios da produção de matéria seca da parte aérea de plantas de alfafa para a fonte ácido bórico revelaram que, aos 150 dias, a produção das doses de 1 e 3 kg ha⁻¹ foi maior em relação as outras doses da mesma ocasião, fato também observado para as doses de 3 kg ha⁻¹ e 6 kg ha⁻¹, respectivamente, aos 180 e 210 dias (Figura 4.b). A dose de 9 kg ha⁻¹ apresentou menor produção para a maioria dos dias de avaliação, principalmente aos 120 e 150 dias.

Quando se usou como fonte de boro a ulexita, houve maior produção de matéria seca nas doses de 3, 6 e 9 kg ha⁻¹ em relação a dose de 1 kg ha⁻¹ ao longo do período avaliado (Figura 4.c). Além disso, de modo geral, observou-se maior produtividade de matéria seca quando se utilizou a ulexita como fonte de boro em relação ao ácido bórico nas doses de 3, 6 e 9 kg ha⁻¹ (Figura 5).



Figura 4 – Perfis médios ± erro padrão da produção de matéria seca da parte aérea (kg ha⁻¹) de plantas de alfafa por doses de boro (0, 1, 3, 6 e 9 kg ha⁻¹) ao longo dos dias de avaliação. Fontes: (a) controle, (b) ácido bórico (H₃BO₃), (c) ulexita



Figura 5 – Perfis médios ± erro padrão da produção de matéria seca da parte aérea (kg ha⁻¹) de plantas de alfafa por fontes de boro (controle, ácido bórico e ulexita) ao longo dos dias de avaliação. Doses: (a) 0 kg ha⁻¹, (b) 1 kg ha⁻¹, (c) 3 kg ha⁻¹, (d) 6 kg ha⁻¹, 9 kg ha⁻¹

4.2 Diagrama de Hasse e número de graus de liberdade

Embora os modelos lineares mistos sejam uma poderosa ferramenta estatística que permite analisar dados experimentais considerando diferentes estruturas de variâncias e covariâncias, a definição do modelo ainda pode ser vista como uma área relevante de estudo. Neste contexto, o diagrama de Hasse permite melhor visualizar a estrutura experimental (ALCARDE, 2012). Para a construção do diagrama, primeiramente identificou-se a estrutura dos fatores não aleatorizados, conforme a seguir.

4Bloco/9Parcela/6Subparcela,

dos quatro blocos, nove tratamentos e dos seis cortes. Assim, as fontes derivadas dessa estrutura são:

$Bloco + Parcela[Bloco] + Subparcela[Bloco \land Parcela].$

Os diagramas de Hasse com indicação para obtenção e com os números de graus de liberdade envolvendo fatores não aleatorizados são apresentados na Figura 6.



Figura 6 – Diagramas de Hasse com indicação para obtenção dos graus de liberdade (a) e com os valores numéricos dos graus de liberdade (b), envolvendo fatores não aleatorizados do experimento de adubação com boro em alfafa

Para a estrutura dos fatores aleatorizados tem-se:

(2A dicional/(3Fonte*5Dose))*6Corte,

dos dois níveis do fator adicional, três níveis de fonte e cinco de dose (originais acrescidos de um nível referente ao tratamento adicional) e dos seis cortes. As fontes derivadas dessa estrutura são:

Os diagramas de Hasse com indicação para obtenção e com os números de graus de liberdade envolvendo fatores aleatorizados são apresentados, respectivamente, nas Figuras 7 e 8.



Figura 7 – Diagrama de Hasse com indicação para obtenção dos graus de liberdade envolvendo fatores aleatorizados do experimento de adubação com boro em alfafa



Figura 8 – Diagrama de Hasse com os valores numéricos dos graus de liberdade envolvendo fatores aleatorizados do experimento de adubação com boro em alfafa

A utilização do diagrama de Hasse foi fundamental para a construção do modelo misto maximal (34), auxiliando na compreensão e visualização simbólica do relacionamento de todos os fatores presentes no estudo, permitindo a decomposição das fontes de variação e de seus graus de liberdade, garantindo que todos os testes fossem realizados corretamente (Figuras 6 e 8, Tabela 2).

O modelo misto maximal (34), proposto com o auxílio do diagrama de Hasse e da estratégia *top-down*, tem a vantagem de incorporar todos os parâmetros associados à estrutura do delineamento experimental. Além disso, por conter todos os termos e interações de efeitos fixos possíveis, retira toda forma de variação sistemática presente nos dados. Tal estratégia também foi utilizada por Sercundes (2014) ao definir, no início dos estudos de avaliação do conforto animal em aviários, um modelo linear misto maximal para os efeitos fixos. Porém, diferentemente, o autor considera a estrutura do esquema fatorial de tratamento do experimento em cada termo de um modelo polinomial proposto.

Com relação à inclusão de uma possível parte aleatória no modelo, com base no comportamento dos perfis individuais da produção de matéria seca da parte aérea de plantas de alfafa apresentados na Figura 2 da subseção 4.1, adicionaram-se efeitos aleatórios associados individualmente a cada unidade experimental, além do efeito aleatório incorporado ao bloco. A necessidade e composição de efeitos aleatórios incluídos

Fontes de Variação	gl
Bloco	3
Parcela[Bloco]	32
Adicional	1
Fonte[Adicional]	1
Dose[Adicional]	3
$Fonte \# Dose \ [Adicional]$	3
Residuo(a)	24
$Subparcela[Bloco^{Parcela}]$	180
Corte	5
A dicional # Corte	5
$Fonte[Adicional] \ \# \ Corte$	5
$Dose[Adicional] \ \# \ Corte$	15
$Fonte \# Dose \ [Adicional] \ \# \ Corte$	15
Residuo(b)	135
TOTAL	215

Tabela 2 – Decomposição dos números de graus de liberdade para o experimento de adubação com boro em alfafa

no modelo são discutidas na subseção 4.3.

4.3 Seleção dos efeitos aleatórios

As estimativas das variâncias dos efeitos aleatórios do modelo linear misto maximal (34) associadas, respectivamente, ao bloco e à unidade experimental foram: $\hat{\sigma}_{b_j}^2 = 1471,86 \ e \ \hat{\sigma}_{b_{fdj(a)}}^2 = 12924.$

De acordo com a significância das estatísticas do teste de razão de verossimilhanças (TRV - Tabela 3), as hipóteses 3.2.1 e 3.2.2 foram rejeitadas, indicando que os efeitos aleatórios associados ao bloco e à unidade experimental devem permanecer no modelo (34) e em todos os modelos subsequentes a ele. Além disso, os menores valores de AIC, AICc e BIC também indicam que o modelo (34) é o mais recomendado em relação à estrutura da parte aleatória. Embora o efeito de bloco tenha sido definido como aleatório em função dos objetivos e características do experimento, o teste para a hipótese 3.2.2, envolvendo $\sigma_{b_j}^2 = 0$, foi realizado apenas para evidenciar que o efeito aleatório para esse fator foi definido corretamente.

Modelo	Efeito aleatório	AIC	AICc	BIC	Modelos comparados	Hipótese	TRV	Valor-p
	presente na				(Referência vs Aninhado)			
	matriz \boldsymbol{D}							
(34)	$\sigma_{b_j}^2, \sigma_{b_{fdj(a)}}^2$	$2255,\!1$	$2255,\!3$	$2253,\!3$				
(35)	$\sigma_{b_i}^2$	$2273,\!3$	$2273,\!4$	$2272,\!1$	(34) vs (35)	Hipótese 3.2.1	$20,2^{*}$	< 0,0001
(36)	Nenhum	$2276,\! 6$	$2276,\! 6$	$2278,\!2$	(34) vs (36)	Hipótese 3.2.2	$25,5^{*}$	< 0,0001

 $\begin{array}{l} {\rm Tabela} \; 3 - {\rm Resultados \; dos \; testes \; de \; hipóteses \; envolvendo \; modelos \; com \; diferentes \; partes \; aleatórias \; ajustados \; na \; análise \; da \; matéria \; seca \; para \; o \; experimento \; de \; adubação \; com \; boro \\ \end{array}$

* Significativo ao nível $\alpha=0,05$

4.4 Seleção da estrutura de covariâncias para os resíduos

Os resultados dos testes envolvendo modelos com diferentes estruturas de correlação para a matriz de erros intra-unidades experimentais são apresentados na Tabela 4. O modelo (34) selecionado na segunda etapa da estratégia *top-down* foi comparado a cada uma das estruturas de correlação incorporadas. Considerou-se primeiramente o resultado do teste da razão de verossimilhanças (TRV), significativo apenas para as seguintes estruturas incorporadas: auto-regressiva de primeira ordem com heterogêneidade de variâncias - ARH(1), componentes de variância com heterogêneidade - CSH, componentes de variância com heterogeneidade - CSH, componentes de variância com heterogeneidade - UN(1) e não estruturada - UN. Esse resultado indica que a incorporação dessas estruturas foi importante para os modelos (34)+ARH(1), (34)+CSH, (34)+UN(1) e (34)+UN, diferenciando-os do modelo (34) pressuposto apenas com uma estrutura de correlação simples, isto é, com independência e homogeneidade de variâncias entre os componentes, o que não é realístico para muitos dados longitudinais (LITTELL; PENDERGAST; NATARAJAN, 2000).

Pelos critérios AIC e AICc, existem evidências de que o modelo com estrutura UN(1) seja o mais apropriado dentre aqueles testados (Tabela 4). Com o teste da razão de verossimilhanças (TRV), comparou-se os modelos (34)+ARH(1) vs (34)+UN(1), (34)+CSH vs (34)+UN(1) e (34)+UN vs (34)+UN(1). Observa-se pela Tabela 4, que não houveram resultados significativos nos dois primeiros testes, indicando, também, que o modelo com estrutura UN(1) é o mais apropriado. Apesar do modelo com estrutura UN ter apresentado menor valor de BIC (2231,5) dentre todos os modelos testados e ter apresentado TRV significativo no teste (34) + UN vs (34) + UN(1), ele não foi selecionado por não ser um modelo parcimonioso e por ter AICc (2253,6) superior ao do modelo (34) + UN(1). Segundo Pinheiro e Bates (2000), se existirem poucas observações por unidade experimental, como no caso deste estudo, a estrutura de correlação geral é mais útil como ferramenta exploratória para determinar estruturas mais simples, que conduzam a modelos mais parcimoniosos. Nesse sentido, a estrutura UN(1) é a melhor opção para a matriz de erros intra-unidades experimentais (\mathbf{R}_i) associada ao modelo linear misto maximal (34), que além dos resultados favoráveis, possui apenas seis parâmetros capazes de modelar a heterogeneidade de variâncias presente nos diferentes cortes da cultura da alfafa.

 $\begin{array}{l} {\rm Tabela}\; 4 - {\rm S}{\rm (intese}\; {\rm das\; estat {\rm isticas\; obtidas\; nos\; testes\; envolvendo\; o\; modelo linear\; misto\; (34)\; com diferentes \\ {\rm estruturas\; de\; correlação\; para\; a\; matriz\; de\; erros\; intra-unidades\; experimentais\; ajustados\; na \\ {\rm análise\; da\; matéria\; seca\; para\; o\; experimento\; de\; adubação\; com boro } \end{array}$

Modelo	AIC	AICc	BIC	Modelos comparados	TRV	Valor-p
				(Referência vs Aninhado)		
(34)	2255,1	$2255,\!3$	$2253,\!3$			
(34) + CS	2257,1	$2257,\!4$	2254,7	(34) + SC vs (34)	$1,\!0505{\rm e}^{-10}$	0,99999
(34) + AR(1)	$2257,\!1$	$2257,\!4$	2254,7	(34) + AR(1) vs (34)	0,0135	0,90769
(34) + ARMA(1,1)	$2256,\!8$	$2257,\!0$	$2254,\!3$	(34) + ARMA(1,1) vs (34)	0,3560	0,83694
(34) + ARH(1)	$2244,\! 6$	$2245,\!7$	2239,0	(34) + ARH(1) vs (34)	$22,\!5683^*$	0,00095
(34) + CSH	2244,5	$2245,\!7$	$2253,\!3$	(34) + CSH vs (34)	$22,\!6087^*$	0,00094
(34) + UN(1)	$2243,\! 0$	$2243,\!9$	2238,1	(34) + UN(1) vs (34)	$22,\!1423^*$	0,00049
(34) + UN	$2245,\! 6$	$2253,\!6$	$2231,\!5$	(34) + UN vs (34)	$49,\!5376^{*}$	0,00026
				(34) + ARH(1) vs $(34) + UN(1)$	0,4307	0,51199
				(34) + CSH vs $(34) + UN(1)$	0,4725	0,49299
				(34) + UN vs $(34) + UN(1)$	$27,\!3919^*$	0,02571

* Significativo ao nível $\alpha=0,05$

4.5 Teste para efeitos fixos

A Tabela 5 apresenta os resultados do teste Wald-F para os efeitos fixos do modelo (34) + UN(1), onde foi verificada a importância de cada um dos termos fixos e suas interações. Como as interações com o fator *Corte* foram significativas, partiu-se para o estudo de regressão.

Fontes de Variação	GL num $^{(1)}$	GL den $^{(2)}$	Wald- F	Valor- p
Adicional	1	24	23,75 *	< 0,0001
Fonte[Adicional]	1	24	66,45 *	< 0,0001
Dose[Adicional]	3	24	8,34 *	0,0006
$Fonte \# Dose \; [Adicional]$	3	24	7,26 *	0,0013
Corte	5	135	$147,\!82$ *	< 0,0001
A dicional # Corte	5	135	4,73 *	0,0005
$Fonte[Adicional] \ \# \ Corte$	5	135	$17,\!69$ *	< 0,0001
$Dose[Adicional] \ \# \ Corte$	15	135	$2,\!52$ *	0,0025
$Fonte \# Dose \; [Adicional] \; \# \; Corte$	15	135	$5,\!48$ *	< 0,0001

Tabela 5 – Teste Wald-F para os efeitos fixos do modelo linear misto (34) com estrutura de componentes de variância com heterogeneidade ("Banded Main Diagonal - UN(1)") para a matriz de erros intra-unidades experimentais (\mathbf{R}_i)

* Significativo ao nível $\alpha=0,05$

⁽¹⁾ Número de graus de liberdade do numerador do teste Wald-F.

 $^{(2)}$ Número de graus de liberdade do denominador do teste Wald-F.

4.6 Estudo de regressão

O estudo de regressão foi realizado com base nos perfis médios da produção de matéria seca da parte aérea de plantas de alfafa para cada um dos tratamentos ao longo dos dias de avaliação apresentados na Figura 3. Em princípio, por facilidade, ajustou-se um modelo de segundo grau aos dados. Porém, os resíduos condicionais associados a este modelo apresentaram padrões de periodicidade ao longo do fator longitudinal. Seguindo a metodologia proposta por Pinheiro e Bates (2000), componentes trigonométricos foram incorporados na parte fixa do modelo para aumentar a qualidade do ajuste. Porém, os resultados não foram satisfatórios e por isso foram omitidos.

A Tabela 6 apresenta o teste Wald-F para os efeitos fixos do modelo de regressão (37) com estrutura de componentes de variância ("Variance Components - VC") para a matriz de erros intra-unidades experimentais (\mathbf{R}_i). Pode-se observar significância para todos os componentes de regressão, inclusive para o de quinto grau. Dessa forma, manteve-se o modelo (37), no qual foram incorporadas diferentes estruturas para a matriz de erros intra-unidades experimentais (\mathbf{R}_i).

Os resultados dos testes envolvendo modelos com diferentes estruturas de correlação para a matriz de erros intra-unidades experimentais encontram-se apresentados na Tabela 7. O modelo (37) proposto na abordagem quantitativa (estudo de regressão)

Fontes de Variação	GL num $^{(1)}$	GL den $^{(2)}$	Wald- F	Valor-p
Adicional	1	24	24,91 *	< 0,0001
Fonte[Adicional]	1	24	69,69 *	$<\!0,\!0001$
Dose[Adicional]	3	24	8,75 *	0,0004
$Fonte \# Dose \ [Adicional]$	3	24	$7,\!61$ *	0,0010
Regressão linear	9	135	12,18 *	$<\!0,\!0001$
Regressão quadrática	9	135	50,99 *	$<\!0,\!0001$
Regressão cúbica	9	135	4,32 *	$<\!0,\!0001$
$Regress$ ão de 4 o grau	9	135	8,58 *	$<\!0,\!0001$
Regressão de 5º grau	9	135	3,01 *	0,0027

Tabela 6 – Teste Wald-F para os efeitos fixos do modelo (37) com estrutura de componentes de variância ("Variance Components - VC") para a matriz de erros intra-unidades experimentais (\mathbf{R}_i)

* Significativo ao nível $\alpha=0,05$

 $^{(1)}$ Número de graus de liberdade do numerador do teste Wald-F.

⁽²⁾ Número de graus de liberdade do denominador do teste Wald-F.

para o fator corte foi comparado a cada uma das estruturas de correlação incorporadas. Todos os resultados foram semelhantes aos obtidos com a abordagem qualitativa expostos na subseção (4.4), indicando que a estrutura UN(1) também é a melhor opção para a matriz de erros intra-unidades experimentais (\mathbf{R}_i).

 $\begin{array}{l} {\rm Tabela} \ 7-{\rm S}({\rm ntese} \ {\rm das} \ {\rm estat}({\rm sticas} \ {\rm obt}({\rm das} \ {\rm nos} \ {\rm testes} \ {\rm envolvendo} \ {\rm o} \ {\rm modelo} \ {\rm linear} \ {\rm misto} \ (37) \ {\rm com} \ {\rm diferentes} \\ {\rm estruturas} \ {\rm de} \ {\rm correlação} \ {\rm para} \ {\rm a} \ {\rm matriz} \ {\rm de} \ {\rm erros} \ {\rm intra-unidades} \ {\rm experimentais} \ {\rm ajustados} \ {\rm na} \\ {\rm análise} \ {\rm da} \ {\rm matéria} \ {\rm seca} \ {\rm para} \ {\rm o} \ {\rm experimento} \ {\rm de} \ {\rm adubação} \ {\rm com} \ {\rm boro} \\ \end{array}$

Modelo	AIC	AICc	BIC	Modelos comparados	TRV	Valor-p
				(Referência vs Aninhado)		
(37)	2443,2	$2443,\!4$	2441,4			
(37) + CS	$2445,\!2$	$2445,\!5$	$2442,\!8$	(37) + SC vs (37)	$5{,}39\mathrm{e}^{-7}$	0,99941
(37) + AR(1)	$2445,\!2$	$2445,\!5$	$2442,\!8$	(37) + AR(1) vs (37)	0,0134	0,90770
(37) + ARMA(1,1)	2444,9	$2445,\!1$	2442,4	(37) + ARMA(1,1) vs (37)	0,3560	$0,\!83694$
(37) + ARH(1)	2432,7	$2433,\!8$	$2427,\!1$	(37) + ARH(1) vs (37)	$22,\!5683^*$	0,00095
(37) + CSH	$2432,\! 6$	$2433,\!8$	$2427,\!1$	(37) + CSH vs (37)	$22,\!6087^*$	0,00094
(37) + UN(1)	$2431,\! 1$	$2432,\!0$	$2426,\!2$	(37) + UN(1) vs (37)	$22,\!1516^*$	0,00049
(37) + UN	2433,7	$2441,\!7$	$2419,\! 6$	(37) + UN vs (37)	$49,\!5376^*$	0,00026
				(37) + ARH(1) vs $(37) + UN(1)$	0,4204	0,51694
				(37) + CSH vs $(37) + UN(1)$	0,4607	0,49762
				(37) + UN vs (37) + UN(1)	$27,\!3919^*$	0,02571

* Significativo ao nível $\alpha=0,05$

O Teste Wald-F para os efeitos fixos do modelo (37) com estrutura de componentes de variância com heterogeneidade ("Banded Main Diagonal - UN(1)") para a matriz de erros intra-unidades experimentais (\mathbf{R}_i) é apresentado na Tabela 8. Para cada componente de regressão existem nove graus de liberdade associados ao numerador do teste Wald-F. Deste modo, existe un parâmetro linear para cada un dos nove tratamentos testados, um parâmetro quadrático para cada um dos nove tratamentos testados e assim sucessivamente até o componente de quinto grau. O teste Wald-F para o componente de regressão linear testa se, conjuntamente, os nove coeficientes de regressão associados a cada um dos nove tratamentos são iguais a zero, dado que todos os outros fatores estão no modelo. O mesmo teste, mudando os parâmetros de regressão, é realizado para os componentes de segundo, terceiro, quarto e quinto grau. O teste Wald-Ffoi significativo para todos os componentes de regressão, indicando que todos eles têm influência no comportamento dos tratamentos ao longo do tempo (BARBIN, 2013). Pelo fator longitudinal (tempo) estar ligado a épocas do ciclo da cultura, que são sujeitas a grandes variações e, pela significância do teste Wald-F, optou-se por manter todos os componentes de regressão do modelo (37) + UN(1).

Tabela 8 – Teste Wald-F para os efeitos fixos do modelo (37) com estrutura de componentes de variância com heterogeneidade ("Banded Main Diagonal - UN(1)") para a matriz de erros intra-unidades experimentais (\mathbf{R}_i)

Fontes de Variação	GL num $^{(1)}$	GL den $^{(2)}$	Wald- F	Valor- p
Adicional	1	24	23,75 *	< 0,0001
Fonte[Adicional]	1	24	66,45 *	< 0,0001
Dose[Adicional]	3	24	8,34 *	0,0006
$Fonte \# Dose \; [Adicional]$	3	24	7,26 *	0,0013
Regressão linear	9	135	$17,\!43$ *	< 0,0001
Regressão quadrática	9	135	$50,\!50$ *	< 0,0001
Regressão cúbica	9	135	$5{,}09$ *	< 0,0001
Regressão de 4º grau	9	135	$9{,}02$ *	< 0,0001
Regressão de 5º grau	9	135	2,16 *	0,0289

* Significativo ao nível $\alpha=0,05$

 $^{(1)}$ Número de graus de liberdade do numerador do teste Wald-F.

⁽²⁾ Número de graus de liberdade do denominador do teste Wald-F.

4.6.1 Diagnósticos

Nesta seção é realizada uma avaliação gráfica do diagnóstico do modelo de estudo de regressão (37) + UN(1). Os resíduos condicionais estudentizados em relação aos valores ajustados (Figura 9.a) e os resíduos ao longo do fator longitudinal (Figura 9.b) para o modelo escolhido apresentam um comportamento aleatório em torno de zero. Na Figura 9.b não se observam padrões de periodicidade, indicando que o ajuste foi adequado. As poucas observações atípicas foram investigadas e, por fazerem sentido biológico e não influenciarem os resultados, foram mantidas no conjunto de dados.

O gráfico de quantil-quantil para os resíduos condicionais estudentizados (Figura 9.c) indica que os mesmos têm distribuição normal. A qualidade do ajuste do modelo (37) + UN(1) é aceitável diante da avaliação da Figura 9 (d).



Figura 9 – Diagnóstico para o modelo (37)+UN(1): (a) Resíduos condicionais estudentizados em relação ao valores ajustados; (b) Resíduos condicionais estudentizados em função dos dias de avaliação; (c) Gráfico de quantil-quantil para os resíduos condicionais estudentizados; (d) Valores ajustados vs observados

4.7 Comparação de tratamentos

A Figura 10 apresenta os perfis médios (fonte dentro de dose) construídos a partir dos valores preditos pelo modelo (37)+UN(1) nos dias avaliados com intervalos de confiança de 95% calculados com base na parte fixa do modelo. Na Figura 10 (a) não se observam diferenças significativas entre os tratamentos avaliados (controle, ácido bórico e ulexita na dose 1 kg ha⁻¹). O tratamento composto pela combinação da fonte ulexita com a dose 3 kg ha⁻¹ apresenta diferença significativa nos primeiros dias avaliados quando comparados aos tratamentos controle e o que envolve ácido bórico (Figura 10.b). Esse mesmo comportamento com relação a fonte ulexita, mas com as doses é 6 kg ha⁻¹ e 9 kg ha⁻¹ é observado, respectivamente, nas Figuras 10 (c) e 10 (d), sendo que na dose 9 kg ha⁻¹ houve diferença significativa de acordo com os intervalos de confiança para todos os tempos avaliados, exceto aos 210 dias.



Figura 10 – Perfis médios construídos a partir dos valores preditos pelo modelo (37)+UN(1) nos dias avaliados com intervalos de confiança de 95% calculados com base na parte fixa do modelo. Tratamentos (fonte.dose em kg ha⁻¹): (a) 1=Controle.0, 2=H₃BO₃.1, 6=Ulexita.1; (b) 1=Controle.0, 3=H₃BO₃.3, 7=Ulexita.3; (c) 1=Controle.0, 4=H₃BO₃.6, 8=Ulexita.6; (d) 1=Controle.0, 5=H₃BO₃.9, 9=Ulexita.9

A semelhança do tratamento controle com a dose 9 kg ha⁻¹ de boro aplicado na forma de ácido bórico (Figura 10.d) possivelmente se deva à alta solubilidade deste produto, tornando-o prontamente disponível para a planta, que por sua vez não atinge uma produtividade esperada, visto que segundo Malavolta (2006), existe uma faixa estreita entre a deficiência e toxidez do nutriente para a planta. Em relação à ulexita, mesmo aplicando a mesma dose, apresenta baixa solubilidade, disponibilizando o nutriente ao longo dos dias de avaliação, fato este não observado com ácido bórico, que apresentou, no geral, a menor produção de matéria seca da parte aérea de plantas de alfafa em praticamente todas as dosagens, fato este, mais nítido nas doses 3, 6 e 9 kg ha⁻¹ (Figuras 10 e 11).



Figura 11 – Perfis médios construídos a partir dos valores preditos pelo modelo (37)+UN(1) nos dias avaliados com intervalos de confiança de 95% calculados com base na parte fixa do modelo. Tratamentos (fonte.dose em kg ha⁻¹): (a) 1=Controle.0, 2=H₃BO₃.1, 3=H₃BO₃.3, 4=H₃BO₃.6, $5=H_3BO_3.9$; (b) 1=Controle.0, 6=Ulexita.1, 7=Ulexita.3, 8=Ulexita.6, 9=Ulexita.9

5 CONCLUSÃO

O uso de modelos lineares mistos para dados longitudinais em ensaio fatorial com tratamento adicional mostrou-se eficiente no estudo da produtividade de matéria seca da parte aérea de plantas de alfafa.

A utilização do diagrama de Hasse foi fundamental para a compreensão e visualização simbólica do relacionamento de todos os fatores presentes no estudo, permitindo a decomposição das fontes de variação e de seus graus de liberdade, facilitando a especificação do modelo e garantindo que todos os testes estatísticos fossem realizados corretamente.

A utilização da estratégia *top-down* também auxiliou na especificação do modelo e permitiu a inclusão de efeito aleatório associado ao intercepto, que foi essencial para a modelagem do comportamento de cada unidade experimental. A estrutura de componentes de variância com heterogeneidade, incorporada aos resíduos, foi capaz de modelar eficientemente a heterogeneidade de variâncias presente nos diferentes cortes da cultura da alfafa.

O estudo de regressão permitiu avaliar a produtividade de matéria seca da parte aérea da planta (kg ha⁻¹) de cortes consecutivos da cultura da alfafa, envolvendo a comparação de adubações com diferentes fontes e doses de boro.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. Automatic Control, London, v. 19, n. 6, p. 716-723, 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.

ALCARDE, R. Modelos lineares mistos em dados longitudinais com o uso do pacote **ASReml-R**. 2012. 158p. Tese (Doutorado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2012.

BARBIN, D. Planejamento e análise estatística de experimentos agropecuários. Londrina: Mecenas, 2013. 214p.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs: Prentice Hall. 1994. 598p.

BRIEN, C.J. Determining the analysis of variance table. In: BRIEN, C.J. **Statistical Modelling**. Disponível em: http://chris.brien.name/ee2/course/SM06.pdf. Acesso em: 01 mar. 2015.

CNAAN, A.; LAIRD N.M.; SLASOR, P. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. **Statistics in Medicine**, Chichester, v. 16, p. 2349-2380, 1997.

DIGGLE, P.J. An approach to the analysis of repeated measurements. **Biometrics**, Washington, v. 44, n. 4, p. 959-971, Dec. 1988.

FAGERIA, N.K. Níveis adequados e tóxicos de boro na produção de arroz, feijão, milho, soja e trigo em solo de Cerrado. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 4, n. 1, p. 57-62, 2000.

GEISSER, S., GREENHOUSE, S.W. An extension of Box's results on the use of the F distribution in multivariate analysis. The Annals of Mathematical Statistics, Chicago, v. 29, p. 855-891, 1958.

HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, Alexandria, v. 72, p. 320-340, 1977.

HARVILLE, D.A. Extension of the Gauss-Markov theorem to include the estimation of random effects. Annals of Statistics, Hayward, v. 4, p. 384-395, 1976.

HENDERSON, C.R. The Estimation of Genetic Parameters. The Annals of Mathematical Statistics, Chicago, v. 21, p. 309–310, 1950.

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.

HENDERSON, C.R. Applications of Linear Models in Animal Breeding. Guelph: University of Guelph Press, 1984. 462 p.

HIJANO, E.H.; NAVARRO, A. La alfalfa en la Argentina. San Juan: INTA, 1995. 281 p.

HUYNH, H.; FELDT, L.S. Conditions under which mean square rations in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, Boston, v. 65, n. 332, p. 1582-1589, Dez. 1970.

HUYNH, H.; FELDT, L.S. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split plot designs. Journal of Educational and Behavioral Statistics, Boston, v. 1, n. 1, p. 69-82, 1976.

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.

KÉRY, M. Introduction to WinBUGS for ecologists : A Bayesian approach to regression, ANOVA, mixed models and related analyses. San Diego: Academic Press, 2010. 297.

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

LITTELL, R.C.; HENRY, P.R.; AMMERMAN, C.B. Statistical analysis of repeated measures data using SAS procedures. **Journal of Animal Science**, Champaign, v. 76, p. 1216–1231, 1998.

LITTELL, R.C.; PENDERGAST, J.; NATARAJAN, R. Modelling covariance structure in the analysis of repeated measures data. **Statistics in Medicine**, Chichester, n. 19, p. 1793-1819, 2000.

LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D; SCHABENBERGER, O. **SAS system for mixed models.** Cary: Statistical Analysis System Institute, 2006. 834 p.

LOHR, S.L. Hasse Diagram in Statistical Consulting and Theaching. The American Statistician, New York, v. 39, n. 4, p. 376-381, Nov. 1995.

MAUCHLY, J.W. Significance test for sphericity of normal n-variate distribution. The Annals of Mathematical Statistics, New York, v. 11, n. 2, p. 204-209, 1940.

MOREIRA, A.; BERNARDI, A.C.C.; RASSINI. J.B.; FERREIRA, R.P.; OLIVEIRA, P.P.A. Fertilidade do solo e nutrição mineral da alfafa cultivada nos trópicos. São Carlos: Embrapa Pecuária Sudeste, 2006. 37p.

MOREIRA, A.; BERNARDI, A.C.C.; RASSINI. J.B.; FERREIRA, R.P.; OLIVEIRA, P.P.A. Fertilidade do solo e estado nutricional da alfafa cultivada nos trópicos. São Carlos: Embrapa Pecuária Sudeste, 2007. 40p.

MOREIRA, A.; BERNARDI, A.C.C.; RASSINI, J.B. Correção do Solo, Estado Nutricional e Adubação da Alfafa. In: FERREIRA, R.P.; RASSINI, J.B.; RODRIGUES, A.A.; FREITAS, A.R.; CAMARGO, A.C.; MENDONÇA, F.C. **Cultivo e utilização da alfafa nos trópicos**. São Carlos: Embrapa Pecuária Sudeste, 2008. cap.4, p. 95-138.

NUERNBERG, N.J. Técnicas de produção de alfafa. In: PEIXOTO, A.M.; MOURA, J.C.; FARIA, V.P. (Eds.). CONGRESSO BRASILEIRO DE PASTAGEM. Piracicaba. 1999. Anais... Piracicaba: FEALQ, 1986. p.145-160.

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.

PIMENTEL GOMES, F. Curso de Estatística Experimental. Piracicaba: FEALQ, 2009. 451 p.

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

RASSINI, J.B.; FERREIRA, R.D.P.; CAMARGO, A.C.D. Cultivo e estabelecimento da alfafa. In: FERREIRA, R.P.; RASSINI, J.B.; RODRIGUES, A.A.; FREITAS, A.R.; CAMARGO, A.C.; MENDONÇA, F.C. **Cultivo e utilização da alfafa nos trópicos**. São Carlos: Embrapa Pecuária Sudeste, 2008. cap.2, p. 37-52.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2015. Disponível em: http://www.R-project.org. Acesso em: 05 maio 2015.

RAIJ, B. Fertilidade do solo e adubação. Piracicaba: Agronômica Ceres/Potafós, 1991. 343p.

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

SAIBRO, J.C. Produção de alfafa no Rio Grande do Sul. In: SIMPÓSIO SOBRE MANEJO DE PASTAGEM, 7., 1984, Piracicaba. Anais... Piracicaba: FEALQ, 1985. p. 61-106.

SARKAR, D. R Development Core Team. lattice: Lattice Graphics, 2015. Disponível em: <http://cran.r-project.org/web/packages/lattice/index.html> R package version, 0.20-31. Acesso em: 05 abr. 2015.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. Akaike information criterion statistics. Tokyo: KTK Scientific, 1986. 320 p.

SAS INSTITUTE INC. SAS course notes on Mixed Model Analysis Using SAS System. Cary, NC, 2002.

SAS INSTITUTE INC. Statistical analysis system user's guide. Version 9.3. ed. Cary: SAS Institute, USA, 2015.

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

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

SELF, S.G.; LIANG, K.Y. Assymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, New York, v. 82, n. 398, p. 605-610, 1987.

SERCUNDES, R.K. Análise de dados longitudinais uma aplicação na avaliação do conforto animal. 2014. 85 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2014.

SHAPIRO, A. Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis. International Statistical Review, Amsterdam, v. 56, n. 1, p. 49-62, 1988.

SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality. **Biometrika**, London, v. 52, p. 591-611, Dez. 1965.

SILVAPULLE, M.J.; SILVAPULLE, P. A Score Test against One-Sided Alternatives. Journal of the American Statistical Association, Alexandria, v. 90, p. 342-349, 1995.

SINGER, J.M.; NOBRE, J.S.; ROCHA, F.M.M. **Análise de dados longitudinais**. Versão parcial preliminar. São Paulo: USP. Departamento de Estatística. Março, 2012. 225 p.

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

SUGIURA, N. Further analysts of the data by akaike's information criterion and the finite corrections. Communications in Statistics - Theory and Methods, Ontario, v. 7, n. 1, p. 13-26, 1978.

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.

TUKEY, J.W. One degree of freedon for non-additivity. **Biometrics**, London, v. 5, p. 232-242, 1949.

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

VERBEKE, G.; MOLENBERGHS, G. The Use of Score Tests for Inference on Variance Components. **Biometrics**, Washington, v. 59, p. 254–262, 2003.

WANG, J. A SAS Macro to Find the Best Fitted Model for Repeated Measures Using PROC MIXED, 2006. Disponível em:

 $<\!http://www.lexjansen.com/pharmasug/2006/ApplicationsDevelopment/AD18.pdf>. Acesso em: 25 maio 2015.$

WANG, L.A.; GOONEWARDENE, Z. The use of MIXED models in the analysis of animal experiments with repeated measures data. Canadian Journal of Animal Science, Ottawa, v. 84, n. 1, p. 1-11, 2004.

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

WICKHAM, H; CHANG, W. R **Development Core Team. ggplot2:** An implementation of the grammar of graphics, 2015. Disponível em: http://cran.r-project.org/web/packages/ggplot2/index.html. R package version, 1.0.1. Acesso em: 31 mar. 2015.

YASSIN, N.; MORAIS, A. R.; MUNIZ, J. A. Análise de variância em um experimento fatorial de dois fatores com tratamentos adicionais. Ciência e Agrotecnologia, Lavras, Edição especial, p.1541-1547, 2002.

ANEXOS

ANEXO A - Códigos computacionais implementados no programa R (R Development Core Team, 2015) para realização da análise exploratória.

```
### Pacotes
pkg <- c("lattice", "latticeExtra", "gridExtra", "doBy", "multcomp",</pre>
"reshape", "plyr", "nlme", "ggplot2", "hnp")
sapply(pkg, require, character.only=TRUE)
### Leitura de dados
dados60a<-read.table('dados60a.txt',h=T,dec='.')</pre>
dados60a <- transform(dados60a,</pre>
Bloco = factor(Bloco),
Dose = factor(Dose),
TRATAMENTO = factor(TRATAMENTO),
TRAT=factor(interaction(Fonte,Dose)))
str(dados60a)
head(dados60a)
### Criação de groupedData (Pacote nlme)
dados60aGD <- groupedData( MS_kgha ~ Dia | Bloco/UNIDEXP,
data = dados60a,
labels = list(x = "Tempo", y = "Matéria Seca"),
units = list(x = "(dias)", y = "(kg/ha)")
)
### Figura 2
plot(dados60aGD, displayLevel = 2)
### Figura 3
ggplot(data=dados60a, aes(x=Dia,y=MS_kgha,group=TRAT,linetype = TRAT)) +
stat_summary(fun.y=mean, geom='line',lwd=1) +
scale_linetype_manual(values = c("twodash", "dashed", "dotted", "dotdash",
"longdash", "solid", "4C88C488", "F1", "1F"),
name="
                                   Tratamentos") +
scale_x_continuous(name='Tempo (dias)',breaks=c(90,120,150,180,210,240)) +
scale_y_continuous(name=expression("Matéria Seca"~( kg~ha^{-1} ))) +
theme_bw() +
theme(axis.text=element_text(size=14),
axis.title=element_text(size=16, #face="bold"
),
legend.position="right", legend.key=element_blank(),
legend.text = element_text(size = 14),
legend.key.width = unit(2.8, "cm"))
```

```
### Figura 4
col=c("black")
\# pch = c(1,4,15,16,17)
lty = c(3, 2, 1, 4, 5)
cex=1.2
xyplot(MS_kgha ~ Dia|Fonte, groups=Dose, data=dados60a, type=c("a"),
par.settings = list(superpose.symbol = list(cex=cex)),
col=col,
lty = lty,
scales=list(x=list(at=c(90,120,150,180,210,240), cex=cex), y=list(cex=cex)),
         auto.key=list(title="Fonte",space="right",cex.title=1.2),
xlab=list(expression("Tempo (dias)"), cex=cex),
ylab=list(expression("Matéria Seca "~(kg~ha^{-1})), cex=cex),
1wd = 1.5,
key=list(corner=c(0.05,0.85), columns=1,
divide=1.
lines=list(col=col, lty = lty, cex=cex, size=10),
text=list(c("0","1","3","6","9")), cex=cex)
)
### Figura 5
col=c("black")
\# pch = c(1, 4, 16)
lty = c(3, 2, 1)
cex=1.2
xyplot(MS_kgha ~ Dia|Dose, groups=Fonte, data=dados60a, type=c("a"),
index.cond=list(c(4,5,2,3,1)),#this provides the order of the panels
par.settings = list(superpose.symbol = list(cex = cex)),
col=col,
lty = lty,
scales=list(x=list(at=c(90,120,150,180,210,240), cex=cex), y=list(cex=cex)),
         auto.key=list(title="Fonte", space="right", cex.title=1.2),
#
xlab=list(expression("Tempo (dias)"), cex=cex),
ylab=list(expression("Matéria Seca"~(kg~ha^{-1})), cex=cex),
1wd = 1.5,
key=list(corner=c(0.95,0.85), columns=1,
divide=1,
lines=list(col=col, lty = lty, size=8, lwd = 1.5),
text=list(c("Controle","Ácido bórico","Ulexita")), cex=cex)
)
```

ANEXO B - Códigos computacionais implementados no programa SAS versão 9.3 (SAS INSTITUTE INC., 2015) para ajuste dos modelos envolvendo os dados do experimento de adubação com boro em alfafa.

```
*;
title "CONJUNTO DE DADOS";
options nodate nonumber center ps=60 ls=100 formdlim='';
PROC IMPORT OUT= WORK.E1
                       /*ALTERAR*/
DATAFILE= "DIRETORIO\dados60a.xlsx"
DBMS=EXCEL REPLACE;
SHEET="P$"; GETNAMES=YES; MIXED=NO; SCANTEXT=YES; USEDATE=YES; SCANTIME=YES;
RUN;
data a;
set WORK.E1;
Corte = (Dia-60)/30;
CorteF = Corte;
DiaF = Dia;
run;
ODS HTML;
proc print data=a;
run;
                                                              *;
************** MODELO PARA ABORDAGEM QUALITATIVA ********************************
*
                                                              *;
* Com Estrutura Simples;
title "MODELO QUALITATIVO";
PROC MIXED DATA=a
                METHOD=REML;
CLASS Adicional Fonte Dose Corte Bloco;
MODEL MS_kgha= Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte Adicional*Corte Adicional*Fonte*Corte
Adicional*Dose*Corte Adicional*Fonte*Dose*Corte / HTYPE=1;
RANDOM Bloco / TYPE=VC;
        Adicional*Fonte*Dose*Bloco / TYPE=VC;
RANDOM
REPEATED Corte / SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=VC;
RUN;
                                                              *;
```

************** ESTUDO DA PARTE ALEATÓRIA - ABORDAGEN QUALITATIVA ******; * Modelo sem RANDOM Adicional*Fonte*Dose*Bloco / TYPE=VC ; PROC MIXED DATA=a METHOD=REML; CLASS Adicional Fonte Dose Corte Bloco; MODEL MS_kgha= Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte Adicional*Fonte*Dose*Corte / HTYPE=1; RANDOM Bloco / TYPE=VC; REPEATED Corte / SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=VC; RUN: ****** Teste da Razão de Verossimilhanças (TRV) para matriz D ****************; title "trvD1 - BLOCO e UNIDADE vs BLOCO"; data trvD1; TRVstat = 2269.3 - 2249.1;q = 2;pvalue = 0.5*(1 - probchi(TRVstat,q)) + 0.5*(1 - probchi(TRVstat,q+1)); format pvalue 10.8; put TRVstat= q= pvalue=; run; proc print data = trvD1; run; * Modelo sem EFEITOS EM b ; PROC MIXED DATA=a METHOD=REML; CLASS Adicional Fonte Dose Corte Bloco; MODEL MS_kgha= Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte Adicional*Fonte*Dose*Corte / HTYPE=1; REPEATED Corte / SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=VC /*R*/; RUN; ******* Teste da Razão de Verossimilhanças (TRV) para matriz D **********;

*;

```
title "trvD2 - BLOCO e UNIDADE vs SEM EFEITOS EM b";
data trvD2;
TRVstat = 2274.6 - 2249.1;
q = 1;
pvalue = 0.5*(1 - probchi(TRVstat,q)) + 0.5*(1 - probchi(TRVstat,q+2));
format pvalue 10.8;
put TRVstat= q= pvalue=;
run;
proc print data = trvD2;
run;
/* Pelo TRV o modelo NÃO deve ficar sem efeito aleatório. */
                                                               *;
```

```
***** ESTUDO DA MATRIZ DE COVARIÂNCIA (R) PARA ABORDAGEM QUALITATIVA
                                                                     *****:
                                                                          *;
title "ESTUDO DA MATRIZ DE COVARIÂNCIA (R) PARA MODELO QUALITATIVO";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a,
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte
Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte
Adicional*Fonte*Dose*Corte,
ClassXs = Adicional Fonte Dose Corte Bloco,
RepeatedVar = Corte,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = CS,
CovS = VC,
CovStructures = VC|CS|CSH|HF|AR(1)|ARH(1)|ARMA(1,1)|TOEP|TOEPH|
UN|UN(1)|UNR|FA(1)|,
NumRept=6);
title "trvR1 - VC vs CS";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a,
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte
Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte
Adicional*Fonte*Dose*Corte,
ClassXs = Adicional Fonte Dose Corte Bloco,
RepeatedVar = Corte,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = CS,
CovS = VC,
CovStructures = VC|CS|,
NumRept=6);
title "trvR2 - VC vs AR(1)";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a,
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte
Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte
Adicional*Fonte*Dose*Corte,
ClassXs = Adicional Fonte Dose Corte Bloco,
```

```
RepeatedVar = Corte,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = AR(1),
CovS = VC,
CovStructures = VC|AR(1)|,
NumRept=6);
title "trvR10 - UN(1) vs UN";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a.
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose Corte
Adicional*Corte Adicional*Fonte*Corte Adicional*Dose*Corte
Adicional*Fonte*Dose*Corte,
ClassXs = Adicional Fonte Dose Corte Bloco,
RepeatedVar = Corte,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = UN,
CovS = UN(1),
CovStructures = UN(1)|UN|,
NumRept=6);
/* ATENÇÃO: o SAS calcula erroneamente o TRV, por colocar NumParaCovG=23*/
title "trvR10 - UN(1) vs UN por GILSON";
data trvR10;
TRVstat = 2226.98 - 2199.59;
NumParaCovD = 2;
NumParaCovS = 6;
NumParaCovG = 21;
NumParaModCovS = NumParaCovD + NumParaCovS;
NumParaModCovG = NumParaCovD + NumParaCovG;
GL = -(NumParaModCovS-NumParaModCovG);
pvalue = 1 - probchi(TRVstat,GL);
format pvalue 10.8;
put TRVstat= NumParaModCovG= NumParaModCovS= GL= pvalue=;
run;
proc print data = trvR10;
run:
*
                                                                         *;
```

```
*;
title "MODELO DE 5º GRAU";
PROC MIXED DATA = a
METHOD=REML;
CLASS Adicional Fonte Dose CorteF Bloco;
MODEL MS_kgha = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose) / residual
outpm=resi3 HTYPE=1; /* HTYPE=1 => SOMA DE QUADRADOS TIPO 1 (SEQUENCIAL,
dado que Adicional está no modelo, testará fonte dentro do Adicional e isso
para os demais) */
RANDOM Bloco / TYPE=VC;
RANDOM Adicional*Fonte*Dose*Bloco / TYPE=VC;
REPEATED CorteF / SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=VC;
RUN;
*
                                                                    *;
**** ESTUDO DA MATRIZ DE COVARIÂNCIA (R) PARA ABORDAGEM QUANTITATIVA ******;
*
                                                                    *;
title "ESTUDO DA MATRIZ DE COVARIÂNCIA (R) PARA MODELO DE 5º GRAU";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a,
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose),
ClassXs = Adicional Fonte Dose CorteF Bloco,
RepeatedVar = CorteF,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
```

```
CovS = VC,
CovStructures = VC|CS|CSH|HF|AR(1)|ARH(1)|ARMA(1,1)|TOEP|TOEPH|UN|UN(1)
|UNR|FA(1)|,
NumRept=6);
```

CovG = CS,

```
103
```

```
************** Teste da Razão de Verossimilhanças (TRV) para matriz R ****;
title "trvRq1 - VC vs CS por GILSON";
data trvRq1;
TRVstat = 0.000000539;
NumParaCovD = 2;
NumParaCovS = 1;
NumParaCovG = 2;
NumParaModCovS = NumParaCovD + NumParaCovS;
NumParaModCovG = NumParaCovD + NumParaCovG;
GL = -(NumParaModCovS-NumParaModCovG);
pvalue = 1 - probchi(TRVstat,GL);
format pvalue 10.8;
put TRVstat= NumParaModCovG= NumParaModCovS= GL= pvalue=;
run;
proc print data = trvRq1;
run;
title "trvRq2 - VC vs AR(1)";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a.
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose),
ClassXs = Adicional Fonte Dose CorteF Bloco,
RepeatedVar = CorteF,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = AR(1),
CovS = VC,
CovStructures = VC|AR(1)|,
NumRept=6);
.
.
title "trvRq10 - UN(1) vs UN";
%include 'DIRETORIO\rmms_random.sas';
%RMMS (
Dt=a,
Y = MS_kgha,
Xs = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
```

```
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose),
ClassXs = Adicional Fonte Dose CorteF Bloco,
RepeatedVar = CorteF,
Sub = Adicional*Fonte*Dose*Bloco,
rand= Bloco Adicional*Fonte*Dose*Bloco,
CovG = UN,
CovS = UN(1),
CovStructures = UN(1)|UN|,
NumRept=6);
/* ATENÇÃO: o SAS calcula erroneamente o TRV, por colocar NumParaCovG=23*/
title "trvRq10 - UN(1) vs UN por GILSON";
data trvRq10;
TRVstat = 2415.08 - 2387.69;
NumParaCovD = 2;
NumParaCovS = 6;
NumParaCovG = 21;
NumParaModCovS = NumParaCovD + NumParaCovS;
NumParaModCovG = NumParaCovD + NumParaCovG;
GL = -(NumParaModCovS-NumParaModCovG);
pvalue = 1 - probchi(TRVstat,GL);
format pvalue 10.8;
put TRVstat= NumParaModCovG= NumParaModCovS= GL= pvalue=;
run;
proc print data = trvRq10;
run;
                                                                      *;
ods html style=journal;
ods graphics on / border = off;
title;
*title "MODELO DE 5<sup>°</sup> GRAU COM ESTRUTURA UN(1)";
PROC MIXED DATA = a
METHOD=REML
PLOTS=STUDENTPANEL(UNPACK) /* Studentized residuals vs. predicted AND
Histogram of studentized residuals AND Q-Q plot of studentized residuals */
CLASS Adicional Fonte Dose CorteF Bloco;
MODEL MS_kgha = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
```

```
106
```

proc gplot data=resi4p;

```
Corte*Corte*Corte*Corte(Adicional Fonte Dose) /
residual outp=resi4p HTYPE=1;
RANDOM Bloco /
             TYPE=VC;
RANDOM Adicional*Fonte*Dose*Bloco / TYPE=VC;
REPEATED CorteF/ SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=UN(1);
RUN;
*
                                                              *;
*;
/* Define symbol characteristics */
symbol1 color=black value=circle;
/* Define axis characteristics */
axis1 order=(60 to 270 by 30) offset=(0,0)
label=("Tempo (dias)");
axis2 order=(-4 to 4 by 1) offset=(0,0)
label=(angle=90 "Resíduo Condicional Estudentizado");
/* Define legend options */
legend1 position=(top center inside)
label=none
mode=share;
* RESIDUOS ESTUDENTIZADOS POR DIA DO MODELO DE 5º GRAU COM ESTRUTURA UN(1);
title;
*footnote ' ';
proc gplot data=resi4p;
plot StudentResid*Dia /
haxis=axis1
vaxis=axis2
vref=-2 to 2 by 4
lvref=2;
run;
*RESIDUOS DE PEARSON POR DIA DO MODELO DE 5º GRAU COM ESTRUTURA UN(1);
axis21 order=(-4 to 4 by 1) offset=(0,0)
label=(angle=90 "Resíduo Condicional de Pearson");
```

```
plot PearsonResid*Dia /
haxis=axis1
vaxis=axis21
vref=-2 to 2 by 4
lvref=2;
run;
                                                            *;
*;
* GRAFICO DOS VALORES OBSERVADOS X ESTIMADOS PELO MODELO DE 5º GRAU COM
ESTRUTURA UN(1) *;
/* Define the title and footnote */
symbol1 color=black value=circle;
/* Define axis characteristics */
axis3 order=(750 to 3000 by 250)
label=("Valores Observados");
axis4 order=(750 to 3000 by 250)
label=(angle=90 "Valores Ajustados");
/* Create an annotate data set that draws a */
/* reference line at a 45 degree angle.
                                */
data anno;
function='move';
xsys='1'; ysys='1';
x=0; y=0;
output;
function='draw';
xsys='1'; ysys='1';
color='black';
x=100; y=100;
output;
run;
proc gplot data=resi4;
plot MS_kgha*Pred/
                   anno=anno
/*legend=legend1 */
haxis=axis3
vaxis=axis4;
run;
                                                            *;
```
```
108
```

```
*;
title "MODELO DE 5<sup>°</sup> GRAU COM ESTRUTURA UN(1)";
PROC MIXED DATA = a
METHOD=REML:
CLASS Adicional Fonte Dose CorteF Bloco;
MODEL MS_kgha = Adicional Adicional*Fonte Adicional*Dose Adicional*Fonte*Dose
Corte(Adicional Fonte Dose)
Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose)
Corte*Corte*Corte*Corte(Adicional Fonte Dose) /
residual
          outpm=resi5pm HTYPE=1;
RANDOM Bloco /
                TYPE=VC;
RANDOM Adicional*Fonte*Dose*Bloco / TYPE=VC;
REPEATED CorteF/ SUBJECT=Adicional*Fonte*Dose*Bloco TYPE=UN(1);
RUN;
* Figura 10 - Compara fonte dentro de dose;
ods html style=journal;
ods graphics on / border = off;
title:
data resi126;
set resi5pm;
if TRATAMENTO in(3,4,5,7,8,9) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi126;
series x=Dia y=Pred2 / group=TRATAMENTO;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
data resi137;
set resi5pm;
if TRATAMENTO in(2,4,5,6,8,9) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi137;
series x=Dia y=Pred2 / group=TRATAMENTO ;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
```

```
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
data resi148;
set resi5pm;
if TRATAMENTO in(2,3,5,6,7,9) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi148;
series x=Dia y=Pred2 / group=TRATAMENTO ;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
data resi159;
set resi5pm;
if TRATAMENTO in(2,3,4,6,7,8) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi159;
series x=Dia y=Pred2 / group=TRATAMENTO ;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
```

* Figura 11 - Compara dose dentro de fonte;

```
data resi12345;
set resi5pm;
if TRATAMENTO in(6,7,8,9) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi12345;
```

```
series x=Dia y=Pred2 / group=TRATAMENTO ;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
data resi16789;
set resi5pm;
if TRATAMENTO in(2,3,4,5) then delete;
if Bloco ne 1 then delete;
Pred2 = Pred;
run;
proc sgplot data=resi16789;
series x=Dia y=Pred2 / group=TRATAMENTO ;
keylegend/noborder title='Tratamentos: ';
scatter x=Dia y=Pred / yerrorupper=Upper yerrorlower=Lower group=TRATAMENTO;
xaxis label='Tempo (dias)' type=linear values=(90 120 150 180 210 240)
valueshint;
yaxis label='Matéria Seca (kg/ha)' values=(750 to 3250 by 250) ;
run;
                                                                      *;
```

110