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

**Uma abordagem para análise de dados com medidas repetidas
utilizando modelos lineares mistos**

Michele Barbosa

Dissertação apresentada para obtenção do título de Mestre
em Agronomia. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2009

Michele Barbosa
Licenciatura em Matemática

**Uma abordagem para análise de dados com medidas repetidas utilizando modelos
lineares mistos**

Orientador:

Prof. Dr. **CÉSAR GONÇALVES DE LIMA**

Dissertação apresentada para obtenção do título de
Mestre em Agronomia. Área de concentração: Es-
tatística e Experimentação Agronômica

Piracicaba

2009

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

Barbosa, Michele

Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos / Michele Barbosa. - - Piracicaba, 2009.
118 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2009.
Bibliografia.

1. Análise de dados longitudinais 2. Leite - Experimentos 3. Medidas repetidas 4. Modelos lineares 5. Software livre I. Título

CDD 519.535
B238a

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

Dedicatória

Aos meus pais e familiares,
Camila,
João Pedro e
Marcos.

“Aventure-se, pois da mais insignificante pista surgiu toda riqueza que o homem já conheceu” (John Masefield).

AGRADECIMENTOS

Desejo externar os meus agradecimentos aos meus pais Antônio e Marlene e à minha irmã Camila pelo incentivo, valorização de minhas decisões pessoais e profissionais, ajuda em todos os momentos e alegria de tê-los ao meu lado sempre.

Ao meu irmão João Pedro, pelo seu divertimento, seu olhar brilhante e a alma cheia de esperança que me permitiu ter o lado infantil bem conservado.

Ao Marcos, pela compreensão durante a elaboração deste trabalho, por compartilhar pequenas e grandes alegrias e fazer uma grande e maravilhosa diferença. *Appreciate your help!*

Prof. Dr. César Gonçalves Lima, por toda a experiência, serenidade e sabedoria transmitidas durante todas as etapas do trabalho e por toda a amizade e paciência.

Ao Prof. Dr. Gerson Barreto Mourão, pela confiança depositada quando do ingresso como estagiária do programa PAE e a cada um dos alunos da graduação, por todos os ensinamentos, contribuindo para a minha formação pessoal e profissional.

Aos participantes do grupo R_Stat, em especial a Walmes Zeviani e Fábio Mathias Côrrea por todas as contribuições oferecidas, que foram fundamentais para o desenvolvimento da pesquisa e ao grupo GEMMIX, que através das trocas de informações com professores e colegas do departamento proporcionou estímulo constante.

Aos professores do departamento de Ciências Exatas ESALQ/USP com os quais tive o prazer de conviver, muito ou pouco, pelos conhecimentos, pela experiência, pelo gosto pela pesquisa e pelo apoio.

À Simone que me acolheu em sua casa na chegada a Piracicaba, à Divisão de Atendimento a Comunidade-Serviço de Promoção Social Vila Estudantil, em especial a todos os alunos moradores da Vila Estudantil, no período de março de 2007 à agosto de 2008, por criarem espaços de encontro entre semelhanças e diferenças, permitindo mais que estar, viver e se doar.

Aos colegas e amigos que me ajudaram, de diferentes formas, em momentos importantes desta trajetória e compartilharam essa experiência comigo, à turma do doutorado, em especial a Fernanda (Loira linda).

À turma de mestrado, em especial as amigas Claudia e Elizabeth, aos amigos Cássio, Elton e Raphael, pelo companheirismo, convivência e cumplicidade.

Aos funcionários da ESALQ e do Departamento de Ciências Exatas, por toda a presteza em fornecer orientações, contribuindo para a qualidade do curso.

Aos técnicos da FZEA, Fábio e Elisângela, pela colaboração prestada.

À Andrezza Maria Fernandes, por disponibilizar os dados.

À Juliana Almeida de Barros (prima querida), pelas devidas correções.

Aos professores da Unesp de Rio Claro, Henrique Lazari e José Silvio Govone, por todo o incentivo para que eu viesse cursar o mestrado na ESALQ.

À Coordenação de Aperfeiçoamento de Pessoa de Nível Superior - CAPES, pelo auxílio financeiro prestado.

Aos meus amigos de Piracicaba, Andradas e de outras cidades do Brasil e do mundo. A convivência com cada um de vocês foi inesquecível e imprescindível.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	19
2 DESENVOLVIMENTO	23
2.1 Revisão bibliográfica	23
2.1.1 Modelos Lineares Mistos	25
2.1.1.1 Introdução e exemplo	25
2.1.1.2 Especificação do modelo	26
2.1.2 Estruturas das Matrizes de Covariância	28
2.1.3 Estimação	31
2.1.4 Seleção de modelos	33
2.1.4.1 Critérios de informação	34
2.1.4.2 Teste da Razão de Verossimilhança	35
2.1.4.3 Teste de Wald	36
2.1.5 Predição dos efeitos aleatórios	36
2.2 O <i>Software</i> R	37
2.2.1 Estruturas da Matriz Positiva Definida (<i>pdMat</i>)	39
2.2.2 Estruturas de Correlação e Função de Variância	40
2.3 O <i>Software</i> SAS	42
2.3.1 Estruturas da Matriz de Covariâncias	42
3 MATERIAL E MÉTODOS	45
3.1 Material	45
3.2 Métodos	49
4 RESULTADOS E DISCUSSÃO	55
4.1 Concentração de α_{S1} -caseína no leite UAT	55
4.1.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$)	60
4.1.2 Diagnóstico do modelo A7.1	62
4.2 Concentração de β -caseína no leite UAT	64
4.2.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$)	68

4.2.2 Diagnóstico do modelo B9.1	71
4.3 Pesos de frango de corte da linhagem <i>Hubbard</i>	72
4.3.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$) (pesos de frangos de corte)	77
4.3.2 Diagnóstico do modelo F6.1	78
4.4 Algumas comparações entre o R e o <code>proc mixed</code> do SAS	79
5 CONCLUSÕES	83
REFERÊNCIAS	85
ANEXOS	89

RESUMO

Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos

No presente trabalho propôs-se uma abordagem simples visando à escolha de um modelo linear misto a ser ajustado a dados com medidas repetidas. A construção do modelo envolveu a escolha dos efeitos aleatórios, dos efeitos fixos e da estrutura de covariâncias utilizando técnicas gráficas e analíticas. O uso do Teste da Razão de Verossimilhança e dos Critérios de Informação de Akaike - AIC e de Schwarz - BIC pode levar a escolhas diferentes da estrutura de covariâncias, o que pode influenciar os resultados das inferências feitas sobre os parâmetros de efeitos fixos. A abordagem foi aplicada a conjuntos de dados resultantes de estudos agropecuários utilizando o *software* livre R. Foram feitas comparações dos resultados obtidos de modelos implementados com o `proc mixed` do SAS e com a função `lme()` do R, observando as vantagens e restrições destes dois *softwares*.

Palavras-chave: Modelo linear misto; Medidas repetidas; Estrutura de covariância; Informação de Akaike; *Software* R.

ABSTRACT

One approach to analyzing data with repeated measures using linear mixed models

In this present work was proposed a simple approach to know how to choose a linear mixed model that can be adjustable to data with repeated measures. The construction of the model involved the choice of random effects, the fixed effects and covariance structure, using graphical and analytical techniques. The use of the Likelihood Ratio Test and the Akaike Information Criteria - AIC and Schwarz - BIC can lead to different choices of the structure of covariance, which may influence the results of inferences made about the parameters of fixed effects. The approach was applied to data sets that was resulted from farming studies using the software R. Comparisons of the results of models implemented were made with the `proc mixed` of SAS and with the function `lme()` of R, noting the advantages and limitations of these two softwares.

Keywords: Linear mixed model; Repeated measures; Structure of covariance; The Akaike Information; Software R.

LISTA DE FIGURAS

Figura 1 - Perfis individuais de resposta (dados hipotéticos)	25
Figura 2 - Perfis individuais das concentração de α_{S1} -caseína (mg/mL) do leite UAT nas cinco ocasiões de armazenamento, por grupo	46
Figura 3 - Perfis individuais das concentrações de β -caseína (mg/mL) do leite UAT nas cinco ocasiões de armazenamento, por grupo	47
Figura 4 - Perfis médios de resposta da concentração de α_{S1} -caseína e β -caseína do leite UAT por grupo	47
Figura 5 - Perfis individuais de peso corporal dos frangos de corte durante sete semanas de idade	48
Figura 6 - Perfil médio do peso corporal de frangos de corte <i>Hubbard</i> , por sexo e por idade	49
Figura 7 - Retas ajustadas aos perfis individuais de α_{S1} -caseína de leite UAT-modelo A1 .	58
Figura 8 - Retas ajustadas para concentração de α_{S1} -caseína (mg/mL) no leite UAT durante o tempo de estocagem, para cada grupo-modelo A7.1	62
Figura 9 - Diagnóstico do ajuste do modelo A7.1	63
Figura 10 - Distribuição dos resíduos do modelo linear por lote de leite UAT	65
Figura 11 - Intervalos de confiança para intercepto e termo linear da reta para cada lote com $\gamma=95\%$	65
Figura 12 - Relação entre os dados originais, o modelo B1 (pontilhado) e modelo B6.1 (linha contínua) para concentração de β -caseína (mg/mL).	68
Figura 13 - Retas ajustadas para a concentração de β -caseína (mg/mL) no leite UAT durante o tempo de estocagem, para cada grupo	70
Figura 14 - Evolução da concentração de β -caseína (mg/mL) do leite UAT durante o armazenamento para cada grupo, segundo o modelo B9.1	71
Figura 15 - Diagnóstico do ajuste do modelo B9.1 (β -caseína)	72
Figura 16 - Boxplot dos pesos corporais dos frangos de corte, por sexo	73
Figura 17 - Distribuição dos resíduos do modelo F1 por indivíduo	75
Figura 18 - Equação ajustada do modelo F6.1	78
Figura 19 - Diagnóstico do ajuste do modelo F6.1	79

LISTA DE TABELAS

Tabela 1 - Classes de estruturas das matrizes de covariância (pdMat) positivas definidas . .	40
Tabela 2 - Classes das funções de variância (varFunc)	41
Tabela 3 - Classes das estruturas de correlação (corStruct)	41
Tabela 4 - Algumas das estruturas de covariância disponíveis pelo proc mixed do SAS . .	43
Tabela 5 - Médias e erros padrões (e.p.) da concentração de α_{S1} -caseína e β -caseína em mg/mL do leite AUT nos cinco tempos de estocagem por grupo	46
Tabela 6 - Médias e erros padrões (e.p.) dos pesos em gramas, de frangos de corte <i>Hubbard</i> , por sexo e semana de idade	48
Tabela 7 - Valores médios, mínimo e máximo de α_{S1} -caseína (mg/mL) do leite UAT nos cinco tempos de estocagem	55
Tabela 8 - Relação de modelos lineares mistos utilizados na análise de dados de leite UAT em concentração de α_{S1} -caseína (mg/mL)	56
Tabela 9 - Estatística de ajuste dos modelos A1 ao A4 (α_{S1} -caseína)	59
Tabela 10 - Testes de Wald para os parâmetros de efeitos fixos do modelo A3.1	59
Tabela 11 - Estatísticas de ajuste dos modelos A4.1, A5 e A6 (α_{S1} -caseína)	60
Tabela 12 - Níveis descritivos (p-valores) dos testes para efeitos fixos, admitindo diferentes estruturas de covariâncias para matriz $\mathbf{R}_i$	60
Tabela 13 - Estatísticas de ajuste dos modelos A4 e A7 (MV) (α_{S1} -caseína)	61
Tabela 14 - Estimativas dos parâmetros fixos do modelo A7.1	61
Tabela 15 - Valores médios, mínimo e máximo de β -caseína (mg/mL) do leite UAT nos cinco tempos de estocagem	64
Tabela 16 - Relação dos modelos lineares mistos utilizados na análise de dados de concentração de β -caseína (mg/mL) no leite UAT	66
Tabela 17 - Estatísticas de ajuste dos modelos B1 ao B3 (MVR) (β -caseína)	67
Tabela 18 - Estatísticas de teste para os parâmetros de efeitos fixos dos modelos B3.1 ao B5	67
Tabela 19 - Estatísticas de ajuste dos modelos B3.1 ao B6 (β -caseína)	68
Tabela 20 - Estatísticas de ajuste dos modelos B6.1 ao B8 (MVR) (β -caseína)	69
Tabela 21 - Níveis descritivos (p-valores) dos testes para efeitos fixos, admitindo diferentes estruturas de covariâncias para a matriz $\mathbf{R}_i$	69
Tabela 22 - Estimativas dos parâmetros de efeito fixo do modelo B7	69
Tabela 23 - Estatísticas de ajuste dos modelos B7.1 ao B10 (β -caseína)	70

Tabela 24 -Relação dos modelos lineares mistos utilizados na análise dos dados de peso corporal dos frangos de corte	74
Tabela 25 -Estatísticas de ajuste dos modelos de F1 ao F3 (MVR) (peso corporal de frangos)	75
Tabela 26 -Estatísticas dos parâmetros de efeitos fixos dos modelos F2.1, F4 e F5	76
Tabela 27 -Estatísticas de ajuste dos modelos de F2.1, F4 e F5 (MVR) (peso corporal de frangos)	76
Tabela 28 -Estatísticas de ajuste dos modelos de F5 ao F9 (MV) (peso corporal de frangos)	77
Tabela 29 -Resultados dos modelos A1 e A2 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (α_{S1} -caseína)	90
Tabela 30 -Resultados dos modelos A3, A3.1 e A4 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (α_{S1} -caseína)	91
Tabela 31 -Resultados dos modelos A4.1, A5 e A6 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (α_{S1} -caseína)	92
Tabela 32 -Resultados dos modelos A7 e A7.1 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (α_{S1} -caseína)	93
Tabela 33 -Resultados do modelo B1 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (β -caseína)	94
Tabela 34 -Resultados dos modelos B2 e B3 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (β -caseína)	95
Tabela 35 -Resultados dos modelos B3.1 e B4 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (β -caseína)	96
Tabela 36 -Resultados dos modelos B5, B6 e B6.1 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (β -caseína)	97
Tabela 37 -Resultados dos modelos B7 e B8 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (β -caseína)	98
Tabela 38 -Resultados dos modelos F1 e F2 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (peso corporal de frangos)	99
Tabela 39 -Resultados dos modelos F2.1 e F3 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (peso corporal de frangos)	100
Tabela 40 -Resultados dos modelos F4, F5 e F6 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (peso corporal de frangos)	101
Tabela 41 -Resultados dos modelos F6.1 e F7 produzidos pelos <code>proc mixed</code> do SAS e função <code>lme()</code> do R (peso corporal de frangos)	102

Tabela 42 -Resultados dos modelos F8 e F9 produzidos pelos <code>proc mixed</code> do SAS e função	
<code>lme()</code> do R (peso corporal de frangos)	103

1 INTRODUÇÃO

Dados de experimentos em que se tomam medidas repetidas de uma ou mais variáveis respostas em ocasiões sucessivas na mesma unidade experimental, ao longo de um certo intervalo de tempo, são comuns em pesquisas nas áreas médica, biológica, econômica, agropecuária etc.

Considerando que neste tipo de problemas, as medidas repetidas são feitas de modo sistemático em cada unidade experimental, é comum admitir-se correlação não nula entre observações feitas em ocasiões distintas e heterogeneidade de variâncias nas diversas ocasiões. Neste contexto, uma abordagem apropriada à análise deve envolver a especificação de um modelo para os valores médios em cada uma das ocasiões e uma estrutura para a matriz de covariâncias entre as medidas feitas ao longo do tempo.

Segundo Helms (1992), as pesquisas com dados longitudinais têm uma longa história em muitas áreas científicas e a terminologia usada não é padronizada. Os dados provenientes de estudos longitudinais são chamados regulares em relação ao tempo se o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do estudo e de balanceados em relação ao tempo se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais.

A análise estatística desse tipo de dados é feita utilizando-se técnicas multi ou univariadas que, geralmente, são dirigidas para o caso de dados completos e balanceados em relação ao tempo. Como exemplos clássicos dessas técnicas de análise, pode-se citar a análise (multi ou univariada) de perfis, a análise de curvas de crescimento etc.

Rocha (2004) salienta sobre as vantagens em estudos com medidas repetidas, que permitem avaliar mudanças globais na resposta com mais eficiência e requerem um número menor de unidades amostrais relativamente a estudos do tipo transversal (“cross sectional”), que envolvem uma única observação da variável resposta em cada unidade experimental. As maiores desvantagens de estudos com medidas repetidas estão relacionadas aos aspectos técnicos, pois, as análises estatísticas são, em geral, mais difíceis e envolvem um alto custo para garantir a observação das unidades amostrais nos instantes pré-determinados. Uma outra desvantagem está relacionada com a presença de dados incompletos, o que na prática ocorre com frequência.

O uso de Modelos Lineares Mistos na análise de dados com medidas repetidas (LAIRD; WARE, 1982) proporcionou uma maior versatilidade na modelagem da matriz de covariâncias, podendo ser conduzida com dados incompletos, irregulares ou desbalanceados em

relação ao tempo, além de englobar as análises uni e multivariadas. A metodologia de análise utilizando modelos lineares mistos foi rápida e eficientemente implementada em importantes programas estatísticos como o SAS®, S-Plus, BMDP e R dentre outros. Em 1992, foi apresentado o `proc mixed` do SAS e em 1998, a função `lme()` do *software* R.

Crowder e Hand (1990), Vonesh e Chinchilli (1997), Molenberghs e Verbeke (2000), Verbeke e Molenberghs (2005), dentre outros, apresentam uma evolução histórica da utilização de modelos lineares mistos na análise de dados longitudinais.

A estimação dos parâmetros neste tipo de modelos é baseada na verossimilhança dos dados. Quando os dados não são normalmente distribuídos, algumas abordagens envolvendo modelos lineares generalizados para dados com medidas repetidas foram propostas por Venezuela (2003), Verbeke (2005), dentre outros.

A escolha da melhor estrutura da matriz de covariâncias visa obter uma estrutura parcimoniosa, que explique bem a variabilidade dos dados nas diversas ocasiões e a correlação entre essas medidas, com um número pequeno de parâmetros, o que pode melhorar a eficiência das inferências feitas sobre os parâmetros do modelo proposto para os valores médios nas diversas ocasiões.

Basicamente, a escolha da estrutura de covariâncias mais adequada ao conjunto de dados é feita utilizando-se: o Teste da Razão de Verossimilhança Generalizada, que serve para comparar o ajuste de modelos encaixados, e os Critérios de Informação de Akaike - AIC e de Schwarz - BIC, dentre outros, que são baseados nos valores da verossimilhança do modelo e dependem do número de observações e do número de parâmetros do modelo. Existem ainda algumas técnicas gráficas, como os diagramas paralelos de dispersão, que servem para detectar a presença de indivíduos com observações aberrantes, sugerir a escolha da melhor estrutura de variâncias e covariâncias e um modelo de regressão adequado para as curvas médias de respostas.

O presente trabalho tem como objetivo o estudo das diversas ferramentas utilizadas na escolha de estruturas de covariâncias e de modelos de regressão para as respostas médias dos grupos, em dados longitudinais, avaliando a aplicação dos modelos lineares mistos na análise de dados de um trabalho com leite obtido pelo processo UAT (temperatura ultra-alta) de Fernandes (2007), que são caracterizados como não regulares, completos e balanceados em relação ao tempo e os dados de frango de Lima (1988). Todas as análises serão implementadas através do *software* R, que disponibiliza uma ampla variedade de metodologias estatísticas, técnicas gráficas, tendo a vantagem de ser um *software* livre.

Buscou-se ainda comparar as implementações de análises utilizando o `proc mixed`

do SAS[®] e os argumentos da função `lme()` do R, evidenciando as vantagens e restrições desses dois *softwares* estatísticos.

2 DESENVOLVIMENTO

2.1 Revisão bibliográfica

Segundo Diggle (1988) e Crowder e Hand (1990), dentre outros, o termo *medidas repetidas* refere-se àqueles casos em que se observa uma ou mais variáveis respostas repetidamente na mesma unidade experimental.

Nesses estudos as variáveis respostas podem ser contínuas (peso, ganho de peso, consumo, conversão alimentar etc.) ou discretas (contagem de algum evento, presença ou ausência de algum sintoma etc.). As unidades experimentais como indivíduos, plantas, animais, canteiros etc., podem estar classificadas em diferentes grupos, segundo um ou mais fatores (ou tratamentos) como sexo, tipo de ração consumida, densidade de plantio, espaçamento entre linhas de plantio etc.

Lima (1996), assim como Vonesh (1997), Littell et al. (1998), dentre outros, relataram que as técnicas utilizadas para analisar os dados de um experimento com medidas repetidas vão desde a análise de variância uni e multivariada, até a metodologia baseada em modelos lineares mistos, com modelagem da estrutura da matriz de variâncias e covariâncias. Everitt (2004) destacou a importância do uso desses modelos na análise de dados com medidas repetidas.

De acordo com Nemec (1996 apud VIEIRA et al., 2007), em uma análise univariada as observações referentes às medidas repetidas são tratadas separadamente, sendo o tempo incluído como um fator no Modelo de Análise de Variância - ANOVA. Na Análise de Variância Multivariada - MANOVA pode-se representar a dependência entre as medidas repetidas e verificar se quaisquer diferenças ocorreram em indivíduos. Esta análise usa um conjunto de suposições menos restritivas que as do modelo de ANOVA e não requer que a variância das respostas nas diversas ocasiões ou que a correlação entre os pares de medidas repetidas permaneçam constantes ao longo do tempo.

O uso do esquema em parcelas subdivididas no tempo, que corresponde a uma análise univariada dos perfis, nem sempre é recomendado para análise de dados com medidas repetidas, porque pressupõe que a estrutura da matriz de covariância seja do tipo uniforme, (variâncias iguais nas diversas ocasiões e covariâncias iguais entre duas ocasiões quaisquer).

Huynh e Feldt (1970) mostraram que uma condição suficiente e necessária para que as estatísticas dos testes de hipóteses envolvendo comparações intra-indivíduos tenham distribuição F exata, é que a matriz de covariâncias satisfaça à condição de esfericidade ou circularidade. A condição Huynh e Feldt - H-F é equivalente a especificar que as variâncias das diferenças entre pares de erros sejam todas iguais. As variâncias das medidas feitas nas mesmas ocasiões

sendo iguais, é equivalente à simetria composta.

Para testar se a matriz de covariâncias atende à condição H-F, Mauchly (1940) propôs o Teste de Esfericidade (ou Circularidade), que verifica se uma população normal multivariada apresenta variâncias iguais e correlações nulas. Caso satisfaça essa condição, a matriz de covariâncias será chamada de *esférica*.

Se a condição de esfericidade da matriz de covariâncias for satisfeita (ou seja, se o teste de Mauchly resultar não significativo), pode-se analisar os dados utilizando-se a técnica de análise univariada de perfis, admitindo um modelo de parcelas subdivididas no tempo. Se a condição de esfericidade da matriz de covariâncias não for satisfeita (ou seja, se o teste de Mauchly resultar significativo), o ideal é realizar uma análise multivariada de perfis. Uma alternativa univariada consiste em corrigir os graus de liberdade das estatísticas dos testes envolvendo comparações intra-indivíduos, realizando uma análise univariada aproximada. Alguns autores, como Geisser e Greenhouse (1958) e Huynh e Feldt (1976), sugerem tais correções, mesmo que a condição de esfericidade não seja satisfeita.

Perri et al. (1999), afirma que uma abordagem mais atual consiste no uso de modelos lineares mistos, que baseia-se em três aspectos fundamentais: estimação e teste de hipóteses sobre os parâmetros de efeito fixo, predição dos parâmetros de efeito aleatório e estimação dos componentes de variância e que segundo Camarinha Filho (2002), o sucesso do procedimento de modelagem está fortemente associado ao exame dos efeitos aleatórios e à possibilidade de se introduzir, no modelo, estruturas de variâncias e covariâncias.

De acordo com Littell et al. (2000) os modelos lineares mistos foram desenvolvidos por geneticistas para avaliar o potencial genético de touros. Sua aplicação foi disseminada por várias áreas de investigação, dinamizada pela disponibilidade de avanços computacionais. Antes desses avanços as análises de modelos mistos eram executadas adaptando-se métodos para modelos de efeitos fixos. Isso trazia limitações à aplicabilidade porque as estruturas de covariâncias não eram modeladas, como é o caso das análises realizadas pelo `proc glm` no SAS. As versões recentes do SAS incluem o `proc mixed`, que permite a modelagem da estrutura de covariância dos dados e o cálculo de estimativas eficientes dos efeitos fixos e de seus respectivos erros padrões.

Verbeke e Molenbergs (2000) ressaltam que muitas vezes os usuários encontram dificuldades em utilizar a metodologia de maneira eficaz. Cursos e consultorias neste domínio têm sido uma grande procura ao longo dos últimos anos, ilustrando a clara necessidade de produção de recursos materiais para ajuda ao usuário.

2.1.1 Modelos Lineares Mistos

2.1.1.1 Introdução e exemplo

Segundo Pinheiro (1994), os modelos lineares mistos têm sido um tema de crescente interesse em Estatística nos últimos cinquenta anos, pois possibilitam a modelagem de correlação intra-indivíduo, muitas vezes presente em dados agrupados. Observações feitas no mesmo indivíduo não podem ser consideradas não correlacionadas e os modelos lineares mistos constituem uma ferramenta conveniente para modelar essa dependência intra-indivíduos.

Figura 1 - Perfis individuais de resposta (dados hipotéticos)

Considerando um exemplo simples, suponha que os diferentes perfis individuais de resposta sejam do tipo apresentado na Figura 1. Obviamente, um modelo de regressão linear com intercepto e efeito linear do tempo seja adequado para descrever o comportamento das respostas ao longo do tempo de cada indivíduo separadamente. Os perfis de resposta de diferentes indivíduos tendem a ser representados por retas com diferentes interceptos e com diferentes inclinações.

Pode-se assumir que a resposta y_{ij} , medida no tempo t_{ij} satisfaz

$$y_{ij} = \tilde{\beta}_{i0} + \tilde{\beta}_{i1}t_{ij} + \epsilon_{ij} \quad \begin{array}{l} i = 1, \dots, N \\ j = 0, 1, \dots, n_i \end{array} \quad (1)$$

e que os ϵ_{ij} são independentes e identicamente distribuídos com distribuição $N(0, \sigma^2)$.

Como os indivíduos são amostrados aleatoriamente de uma população de indivíduos, é natural supor que os coeficientes de regressão $\tilde{\beta}_i = (\tilde{\beta}_{i0}, \tilde{\beta}_{i1})^\top$ sejam amostrados aleatoriamente de uma população de coeficientes de regressão. Pode-se reformular o modelo (1) como:

$$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{pi} + \epsilon_{ij} \quad (2)$$

onde $\tilde{\beta}_{i0} = \beta_0 + b_{0i}$ e $\tilde{\beta}_{i1} = \beta_1 + b_{1i}$. Assume-se que os efeitos aleatórios $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$ são identicamente distribuídos com distribuição $N(\mathbf{0}, \mathbf{D})$ e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^\top$ e $\mathbf{b}_i$ independentes.

O modelo (2) é um caso especial do modelo linear misto, uma vez que contém efeitos fixos e aleatórios. Modelos similares são utilizados em outras situações onde há estrutura de agrupamento nos dados, por exemplo, utilizando modelos mistos de Poisson (BRESLOW; CLAYTON, 1993).

Há várias abordagens para os modelos lineares mistos como modelo linear clássico, modelo de componentes de variância e também modelos hierárquicos multiníveis (NATIS, 2000), entre outras alternativas de análise de dados com medidas repetidas como modelos lineares generalizados mistos (SILVANO, 2003).

2.1.1.2 Especificação do modelo

Os modelos lineares mistos permitem a utilização de várias estruturas de covariâncias no processo de modelagem, que foram tratados primeiramente por Laird e Ware (1982), Ware (1985), Jennrich e Schluchter (1986) e Diggle et al. (1998), dentre outros. Eles consideraram os efeitos fixos no primeiro estágio, para obtenção da curva polinomial média e no segundo estágio, permitiram diferentes curvas para cada indivíduo.

O modelo linear misto é expresso na seguinte forma matricial:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i \quad i = 1, \dots, N \quad (3)$$

onde $\mathbf{y}_i$ representa um vetor ($n_i \times 1$) de respostas da i -ésima unidade experimental ou indivíduo, $\mathbf{X}_i$ é uma matriz ($n_i \times p$) de especificação (conhecida e de posto completo) dos efeitos fixos, $\boldsymbol{\beta}$ é um vetor ($p \times 1$) de parâmetros (efeitos fixos), $\mathbf{Z}_i$ é uma matriz ($n_i \times q$) de especificação (conhecida e de posto completo) dos efeitos aleatórios, $\mathbf{b}_i$ é um vetor ($q \times 1$) de efeitos aleatórios com vetor de média $\mathbf{0}$ e matriz de covariância $\mathbf{D}$ e $\boldsymbol{\epsilon}_i$ é um vetor ($n_i \times 1$) de erros aleatórios com vetor de média $\mathbf{0}$ e matriz de covariância $cov(\boldsymbol{\epsilon}_i) = \mathbf{R}_i(\boldsymbol{\lambda})$, positiva definida, com $\boldsymbol{\lambda}$ um vetor

de parâmetros desconhecidos. Há N unidades experimentais e n_i observações feita na i -ésima unidade experimental.

As matrizes de especificação $\mathbf{X}_i$ e $\mathbf{Z}_i$ podem ser diferentes e variar entre unidades experimentais, estendendo o modelo para o caso de dados não balanceados em relação ao tempo. As matrizes $\mathbf{Z}_i$ podem conter quaisquer covariáveis que afetem diferentemente as unidades experimentais.

A forma de especificação da matriz $\mathbf{X}_i$ é bastante similar àquela utilizada nos modelos de regressão. Suas colunas podem estar associadas:

- i) aos fatores que definem a estrutura das subpopulações (grupos ou tratamentos);
- ii) ao fator tempo, identificando, por exemplo, a forma da curva a ser ajustada;
- iii) a covariáveis, cujos efeitos na resposta média deseja-se pesquisar.

Estes pressupostos implicam que a matriz de covariância de $\mathbf{y}_i$ é $cov(\mathbf{y}_i)$:

$$cov(\mathbf{y}_i) = \Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i$$

Esse método de estruturar a matriz de covariâncias Σ_i tem como atrativo a possibilidade de:

- i) englobar as abordagens uni e multivariada que são comumente utilizadas na análise de dados longitudinais;
- ii) lidar com dados perdidos, por causa da facilidade de construir a verossimilhança somente dos dados observados;
- iii) usar estruturas relacionadas com séries temporais ou estruturas mais complexas.

Em situações onde o objetivo da análise é ajustar curvas (de crescimento), pode-se dizer que os modelos lineares mistos assumem a existência de curvas subpopulacionais fixadas ($\mathbf{X}_i \boldsymbol{\beta}$) em torno das quais existem variações aleatórias ($\mathbf{Z}_i \mathbf{b}_i$) das curvas individuais e também, que existem variações aleatórias de medidas ($\boldsymbol{\epsilon}_i$) em torno dessas curvas individuais.

O modelo linear misto é geralmente especificado em termos das respostas condicionadas aos efeitos aleatórios, de modo que assumindo $\mathbf{y}_i$ um vetor de medidas repetidas para o i -ésimo indivíduo, satisfaça:

$$\begin{aligned} \mathbf{y}_i | \mathbf{b}_i &\sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \end{aligned}$$

Quando a matriz de covariância do erro aleatório for igual a $\sigma^2 \mathbf{I}$, $\mathbf{I}$, tem-se o modelo conhecido como *modelo com independência condicional*, que reflete a independência e a homocedasticidade das observações intra-indivíduos. Entretanto, as inferências são baseados no modelo marginal, a menos que uma abordagem Bayesiana seja empregada para a análise.

$$\mathbf{y}_i \sim N(\mathbf{X}_i \beta, \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i)$$

2.1.2 Estruturas das Matrizes de Covariância

A escolha da estrutura de covariância afeta as estimativas e os erros padrões de efeitos fixos, diagnósticos e inferências. A escolha depende de informação empírica, da estrutura dos dados e muitas vezes da disponibilidade computacional.

A metodologia de modelos lineares mistos permite a consideração de formas especiais para a matriz de covariância, que buscam representar a variabilidade dos dados da forma mais real possível, ou seja, levam em consideração se os dados são independentes, dependentes, correlacionados etc.

Serão apresentadas, a seguir, algumas estruturas das matrizes $\mathbf{D}$ e $\mathbf{R}_i$ mais utilizadas e que se encontram implementadas em *softwares* estatísticos como SAS e R. Considerando $n_i=4$ ocasiões, tem-se:

1. Componente de Variância (VC)

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

Impõe variâncias iguais nas n_i ocasiões e observações independentes. Envolve um único parâmetro.

2. Simetria Composta (CS)

$$\begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 \end{pmatrix}$$

Impõe variâncias iguais nas n_i ocasiões e mesma covariância entre medidas feitas em ocasiões distintas. Envolve dois parâmetros.

3. Simetria Composta Heterogênea

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

Impõe parâmetros de variâncias diferentes para cada elemento da diagonal principal e raiz quadrada desses parâmetros nos elementos fora da diagonal principal, sendo σ_i^2 o i -ésimo parâmetro da variância e ρ representa a correlação entre as medições, satisfazendo $|\rho| < 1$. Envolve $n_i + 1$ parâmetros.

4. Não Estruturada (UN)

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Impõe variâncias distintas para cada uma das n_i ocasiões e covariâncias diferentes entre medidas feitas em ocasiões distintas. Envolve $n_i(n_i + 1)/2$ parâmetros.

5. Estrutura AR(1)

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

Impõe variâncias iguais nas diversas ocasiões e correlação decrescente com o aumento do intervalo entre as ocasiões. Envolve dois parâmetros.

6. Estrutura ARH(1)

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

É uma generalização da estrutura AR(1), impondo variâncias e covariâncias diferentes. Envolve $n_i + 1$ parâmetros.

7. Estrutura Toeplitz

$$\begin{pmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{pmatrix}$$

É a estrutura de covariâncias de um processo de médias móveis de ordem $q = n_i - 1$ (neste exemplo $q=3$). Envolve n_i parâmetros.

8. Estrutura ARMA(1,1)

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

É a estrutura associada a séries temporais com parâmetro auto-regressivo ρ , componente de médias móveis γ , sendo σ^2 a variância residual. Envolve três parâmetros.

9. Estrutura Ante-Dependência de ordem 1

$$\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ \sigma_1\sigma_3\rho_1\rho_2 & \sigma_2\sigma_3\rho_2 & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \sigma_1\sigma_4\rho_1\rho_2\rho_3 & \sigma_2\sigma_4\rho_2\rho_3 & \sigma_3\sigma_4\rho_3 & \sigma_4^2 \end{pmatrix}$$

Impõe parâmetros de variâncias diferentes para cada elemento da diagonal, sendo os elementos fora da diagonal principal funções de variâncias e do k -ésimo parâmetro de autocorrelação, satisfazendo $|\rho_k| < 1$. Esta estrutura permite que as variâncias sejam diferentes e é aplicável em estudos longitudinais em que as condições de avaliação não são igualmente espaçadas, apresentam heterogeneidade de variância e correlação serial. Envolve $2n_i - 1$ parâmetros.

10. Estrutura Toeplitz Heterogênea

$$\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ \sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_1\sigma_4\rho_3 & \sigma_2\sigma_4\rho_2 & \sigma_3\sigma_4\rho_1 & \sigma_4^2 \end{pmatrix}$$

É uma estrutura associada a dados de séries temporais igualmente espaçados, com parâmetros de variâncias diferentes para cada elemento da diagonal, sendo os elementos fora da diagonal funções de variâncias e do k -ésimo parâmetro de autocorrelação $|\rho_k| < 1$.

11. Huynh-Feldt (HF)

$$\begin{pmatrix} \sigma_1^2 & \frac{(\sigma_1^2 + \sigma_2^2)}{2} - \lambda & \frac{(\sigma_1^2 + \sigma_3^2)}{2} - \lambda & \frac{(\sigma_1^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_2^2 + \sigma_1^2)}{2} - \lambda & \sigma_2^2 & \frac{(\sigma_2^2 + \sigma_3^2)}{2} - \lambda & \frac{(\sigma_2^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_3^2 + \sigma_1^2)}{2} - \lambda & \frac{(\sigma_3^2 + \sigma_2^2)}{2} - \lambda & \sigma_3^2 & \frac{(\sigma_3^2 + \sigma_4^2)}{2} - \lambda \\ \frac{(\sigma_4^2 + \sigma_1^2)}{2} - \lambda & \frac{(\sigma_4^2 + \sigma_2^2)}{2} - \lambda & \frac{(\sigma_4^2 + \sigma_3^2)}{2} - \lambda & \sigma_4^2 \end{pmatrix}$$

É a estrutura que impõe variâncias diferentes nas diversas ocasiões e covariâncias calculadas como a média aritmética entre as variâncias e subtraindo λ , onde λ é a diferença entre a média das variâncias e a média das covariâncias. Envolve $n_i + 1$ parâmetros.

Littell et al. (2000) sugerem o ajuste de um modelo inicial saturado tanto para os efeitos fixos quanto para a estrutura de covariância e o ajuste subsequente de estruturas de covariância mais parcimoniosas tais como uniforme, uniforme heterogênea, auto-regressiva etc. Sugerem também que a comparação dessas estruturas de covariância seja feita por meio dos critérios de informação AIC e BIC.

2.1.3 Estimação

Segundo Pinheiro e Bates (2000), dentre os métodos existentes para estimar os parâmetros do modelo (3), os mais comumente usados são o Método da Máxima Verossimilhança - MV e da Máxima Verossimilhança Restrita - MVR, seguido de uma decomposição ortogonal-triangular, afim de facilitar as representações computacionais.

Verbeke e Molenbergs (2000) apresentam a forma clássica de estimação baseada na maximização da função verossimilhança marginal:

$$L_{MV}(\boldsymbol{\theta}) = \prod_{i=1}^N \left\{ (2\pi)^{-\frac{n_i}{2}} |\boldsymbol{\Sigma}_i(\boldsymbol{\alpha})|^{-\frac{1}{2}} \times \exp \left(-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\alpha}) (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right) \right\} \quad (4)$$

no qual $\boldsymbol{\alpha}$ denota o vetor com todos os parâmetros de variâncias e covariâncias (usualmente chamados de componentes de variância) encontrado em $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i$, que consiste de $n_i(n_i + 1)/2$ elementos diferentes em $\mathbf{D}$ e de todos os parâmetros em $\mathbf{R}_i$. Seja $\boldsymbol{\theta}$ o vetor de parâmetros para o modelo marginal $\mathbf{y}_i$.

O estimador de máxima verossimilhança de β , obtidos a partir de maximização de (4), condicionada a α é dada por (LAIRD; WARE, 1982):

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^N \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \Sigma_i^{-1} \mathbf{y}_i$$

Quando α não for conhecida, mas uma estimativa $\hat{\alpha}$ for disponível, pode-se ter $\hat{\Sigma}_i = \Sigma_i(\hat{\alpha})$ e estimar β , utilizando a expressão (4) substituindo Σ_i por $\hat{\Sigma}_i$.

O método de máxima verossimilhança restrita é uma modificação do método de máxima verossimilhança. A dedução do método de verossimilhança restrita é praticamente a mesma, mas ao invés de otimizar diretamente a verossimilhança das observações diretamente, ele otimiza a integral da verossimilhança dos resíduos. Do ponto de vista bayesiano, ignoram qualquer informação prévia sobre os efeitos fixo e utilizam todos os dados para fazer as inferências.

Esta alteração garante, pelo menos nos casos balanceados, que os parâmetros de efeito aleatório sejam estimados sem viés, e por essa razão, o estimador de máxima verossimilhança restrita é, geralmente, preferido em modelos lineares mistos, contendo todas as informações sobre os parâmetros de variância.

Uma definição de estimação de Máxima Verossimilhança Restrita - MVR que fornece uma conveniente forma computacional é

$$L_{MVR}(\theta) = \left| \sum_{i=1}^N \mathbf{X}_i' \Sigma_i^{-1} \mathbf{X}_i \right|^{-\frac{1}{2}} L_{ML}(\theta)$$

As estimações por máxima verossimilhança e máxima verossimilhança restrita têm os mesmos méritos, que se baseiam no princípio que conduz a propriedades úteis de verossimilhança, como a coerência, normalidade assintótica e eficiência. Por outro lado, em modelos lineares mistos, as estimativas por MVR para os componentes de variância são idênticas àquelas obtidas pelo método dos momentos. Exemplos de métodos de estimação também podem ser encontrados em Vonesh e Chinchilli (1997).

Para maximizar o logaritmo da verossimilhança restrita são necessários métodos iterativos, como o método de *Newton-Raphson*, o método *Fisher scoring* e o método *EM* proposto por Laird e Ware (1982). O método de Newton-Raphson com as modificações propostas por Jennrich e Schluchter (1986), é considerado melhor que os demais em relação ao tempo total para atingir a convergência.

No `proc mixed` do SAS e nas funções `lme()` e `gls()` do *software* R estão implementados os métodos de estimação MV e MVR utilizando os algoritmos computacionais *Ridge-stabilized N-R* e *Fisher scoring* (SAS) e *EM algorithm* e *Newton-Raphson* (R).

2.1.4 Seleção de modelos

Selecionar o melhor modelo significa não só selecionar a melhor estrutura para as médias (parte fixa), como também a melhor estrutura de covariâncias. A construção do modelo consta de três etapas: seleção dos efeitos fixos, identificação dos efeitos aleatórios, estimação e comparação de modelos.

Rocha (2004) propõe uma série de técnicas gráficas e analíticas que auxiliam a escolha das matrizes dos modelos lineares mistos, afirmando que para estudos longitudinais, é razoável utilizar informações sobre o comportamento da resposta ao longo das ocasiões de avaliação na modelagem da estrutura de covariância intra-unidades amostrais.

O processo de seleção e avaliação do modelo é uma tarefa simples e alguns métodos de seleção de modelos podem ser utilizados para auxiliar na escolha do modelo que melhor se ajusta aos dados. Técnicas apresentadas por Wolfinger (1993), tornam úteis uma coleção de estruturas de covariância para dados de medidas repetidas, o que amplia significativamente o arsenal de modelos estatísticos disponíveis para explicar a variabilidade dos dados.

A seleção do modelo adequado é realizada frequentemente através do teste da razão de verossimilhança e dos Critérios de Informação de Akaike - AIC e Bayesiano - BIC. O `proc mixed` do SAS permite especificar a estrutura da matriz de covariâncias associada aos efeitos aleatórios, **D**, através do comando `random` e a estrutura da matriz de covariância intra-indivíduos, **R_i**, através do comando `repeated`. Alguns exemplos de estruturas já foram apresentados na seção 2.1.2. O problema que surge é que na prática, a verdadeira estrutura de covariâncias é desconhecida.

Fernandez (2007), apresenta a macro **ALLMIXED2** do SAS que automatiza com eficiência a seleção de modelos lineares mistos. Ela possibilita a escolha das melhores estruturas de covariâncias para dados de medidas repetidas, através de gráficos, do teste da razão de verossimilhança e critérios de informação. No entanto, esta macro só pode ser aplicada no *software* SAS versão 9.13.

Pinheiro e Bates (2000) ressaltaram a crescente popularidade dos modelos lineares mistos, que é explicada pela grande flexibilidade que oferecem na modelagem da correlação intra-indivíduos, pela manipulação de dados balanceados e desbalanceados e pela disponibilidade de *software* confiável e eficiente para o seu ajuste.

Segundo Venables e Ripley (2002), a principal ferramenta para ajuste de modelos lineares mistos é o pacote **nlme3** do *software* R, descrito em Pinheiro e Bates (2000). Pode-se

instalar *library (nlme3)* ou *library (nlme)* para utilizá-lo.

O *software* R apresenta a função denominada `lme()`, que serve para ajustar modelos lineares mistos a dados de medidas repetidas. Esta função está disponível no pacote `nlme`.

Recentemente tem-se desenvolvido o pacote `lme4`, que apresenta a função `lmer()`, que possibilita o ajuste para modelos lineares mistos, modelos não lineares e modelos lineares generalizados. Conforme orientação de Bates (2005, p. 27): “A boa notícia para os usuários da `lme()` é que a função `lmer()` ajusta uma maior gama de modelos, é mais confiável e é mais rápida do que a função `lme()`. A má notícia é que a especificação do modelo foi ligeiramente modificada”.

No entanto, para incluir estruturas de covariância intra-indivíduos, deve ser flexivelmente modelada pela função `lme()` do pacote `nlme`, combinando as estruturas de correlação através de funções da classe (`corStruct`) e funções de variância (`varFunc`). Pinheiro e Bates (1999) utilizam vários exemplos, compondo diferentes estruturas.

2.1.4.1 Critérios de informação

Floriano et al. (2006) afirmaram que muitos métodos já foram desenvolvidos visando facilitar a escolha da estrutura de covariância que melhor explique o comportamento da variabilidade e da correlação entre as medidas repetidas. Os principais critérios de seleção de modelos usados em programas computacionais são o critério de Akaike (*Akaike's Information Criterion*) - AIC e o bayesiano de Schwarz (*Bayesian Information Criterion* - BIC), que são baseados no valor da verossimilhança do modelo e dependem do número de observações e do número de parâmetros do modelo.

Na seleção do modelo mais adequado, é necessário o cálculo do valor de AIC e BIC para cada modelo considerado, obtendo-se uma classificação dos modelos candidatos. A distância entre o verdadeiro modelo e o modelo selecionado pode ser representado pela *informação de Kullback-Leibler* (KULLBACK-LEIBLER, 1978 apud NGO, 2002).

O Critério de Akaike - AIC baseando-se no logaritmo da verossimilhança (MV ou MVR) $L(\boldsymbol{\theta})$ pode ser calculado por:

$$AIC = -2L(\hat{\boldsymbol{\theta}}) + 2d \quad (5)$$

onde, d representa o número total de parâmetros de efeito fixo e aleatório estimado no modelo. Dentre todos os possíveis modelos considerados, o modelo com o menor valor de AIC é considerado o melhor modelo.

O Critério de informação de Schwarz - BIC ou *Schwarz Information Criterion* -

SIC é assim chamado porque Gideon E. Schwarz (1978) apresentou um argumento Bayesiano para prová-lo. O BIC é calculado por:

$$BIC = -2L(\hat{\boldsymbol{\theta}}) + \ln(N)d \quad (6)$$

onde $N = \sum n_i$ (soma do tamanho de todos os vetores $\mathbf{y}_i$). Uma característica do BIC é penalizar os modelos mais complexos, com maior número de parâmetros. Segundo este critério, o melhor dos modelos será o que apresentar o menor BIC.

West et al. (2007) alertam para o fato de que alguns *softwares* calculam os valores de AIC e BIC utilizando fórmulas diferentes, dependendo se a estimação é feita por MV ou MVR.

Bates (2000) ressalta que quando os modelos são ajustados por MVR, os valores de AIC, BIC e log-verossimilhança somente podem ser comparados entre modelos com a mesma estrutura de efeitos fixos. Quando os modelos são ajustados por máxima verossimilhança os valores de AIC e BIC podem ser comparados entre quaisquer modelos ajustados para os mesmos dados. Neste último caso, a qualidade de ajuste pode ser avaliada para diferentes especificações dos efeitos fixos ou diferentes especificações dos efeitos aleatórios ou de ambos.

Gurka (2006) apresenta estudos de simulação sobre o desempenho dos critérios de informação na seleção do melhor modelo, quando se utiliza o método de estimação MVR. Sugere o uso do BIC para tomar uma decisão quando os dois critérios indicam dois modelos diferentes.

2.1.4.2 Teste da Razão de Verossimilhança

A estatística $-2\log\text{Veross}$ é baseado no logaritmo da razão entre as duas verossimilhanças dos modelos mais simples, $l(\tilde{\boldsymbol{\theta}})$, e o modelo mais complexo, $l(\hat{\boldsymbol{\theta}})$

$$-2\log\text{Veross} = -2[\log(l(\tilde{\boldsymbol{\theta}})) - \log(l(\hat{\boldsymbol{\theta}}))] \sim \chi_r^2$$

é assintoticamente distribuída como uma *quiquadrado* com r graus de liberdade e serve para testar a hipótese H_0 : o modelo mais simples é adequado. Quando esta hipótese for rejeitada, conclui-se que o modelo mais complexo (com maior número de parâmetros) é adequado. Se o modelo mais simples for adequado, os valores da função de verossimilhança avaliadas em $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ devem estar próximos, indicando que os dados estão dando suporte ao modelo com menor número de parâmetros.

Segundo Guimarães (1994 apud XAVIER, 2000), embora seja bastante eficiente, a principal desvantagem desse teste é que ele só pode ser usado para comparar dois modelos de cada vez, sendo que um deles deve ser um caso especial do outro.

Pinheiro e Bates (2000) utilizaram o teste da razão de verossimilhança para avaliar a importância dos efeitos aleatórios, ajustando diferentes modelos aninhados em que as estruturas de efeitos aleatórios mudaram da mais complexa para a mais simples.

Stram e Lee (1994 apud PINHEIRO; BATES, 2000) utilizaram resultados de Self e Liang (1987) e afirmaram que os estudos sobre as estruturas de efeitos aleatórios conduzidos desta forma, tendem a ser conservadores, ou seja, que o p-valor calculado a partir da distribuição *quiquadrado* é maior do que deveria ser.

2.1.4.3 Teste de Wald

O teste de Wald serve para avaliar a significância dos efeitos fixos do modelo (3). A estatística de Wald para testar $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$, onde $\mathbf{C}$ ($c \times p$) é uma matriz de constantes conhecidas e de posto completo c ($c \leq p$) é escrita como:

$$\mathbf{Q}_c = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}\widehat{cov}(\hat{\boldsymbol{\beta}})\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}})$$

Onde $\widehat{cov}(\hat{\boldsymbol{\beta}})$ é uma estimativa da matriz de covariâncias de $\hat{\boldsymbol{\beta}}$. Sob H_0 a estatística $\mathbf{Q}_c$ tem distribuição assintótica *quiquadrado* com c graus de liberdade. Dividindo $\mathbf{Q}_c$ por c , obtém-se uma outra estatística que tem distribuição F com c e $p - \text{posto}(\mathbf{X})$ graus de liberdade.

Segundo Verbeke e Molenberghs (2000) o teste Wald não é adequado para uso com modelos lineares mistos, que são especificados condicionalmente aos efeitos aleatórios $\mathbf{b}_i$, isto é $\mathbf{y}_i|\mathbf{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \boldsymbol{\Sigma}_i)$. O teste não leva em conta a estimativa dos parâmetros de efeito aleatório e pode subestimar a variação dos efeitos fixos.

Um teste alternativo para os parâmetros de covariância é o teste-z de Wald. Segundo West et al. (2007), este teste é assintótico e exige que o fator com o qual os efeitos aleatórios estão associados tenha um grande número de níveis. Este teste estatístico também apresenta propriedades desfavoráveis quando testa hipótese sobre parâmetros de covariância, que assumem valores nos limites do seu espaço paramétrico. Devido a estes inconvenientes, ao invés da utilização do teste-z de Wald para os parâmetros de covariância, recomenda-se a utilização do teste da razão de verossimilhança.

2.1.5 Predição dos efeitos aleatórios

No contexto de modelos lineares mistos, pode ser útil prever os valores dos efeitos aleatórios associados a níveis específicos de um fator de efeito aleatório. Para isso, realiza-se a

análise da esperança condicional dos efeitos aleatórios, tendo em conta os valores observados. A fórmula é dada em notação matricial como:

$$\hat{\mathbf{u}}_i = \mathbf{D}\mathbf{Z}_i\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$$

Refere-se a eles como EBLUPs (ou BLUPs empíricos), porque se baseiam na estimativa da matriz de variância e covariância.

Recomenda-se a utilização de gráficos como histogramas, Q-Q plots, afim de obter um bom diagnóstico do modelo.

2.2 O *Software* R

O sucesso da aplicação de qualquer técnica estatística está diretamente relacionado com a disponibilidade de equipamentos computacionais eficientes e o uso de *softwares* simples e confiáveis.

Dentre as ferramentas computacionais estatísticas que permitem realizar análises de dados com medidas repetidas, o uso do *software* gratuito e aberto R- version 2.8.1 (R, 2008) apresenta diversas vantagens. O R é uma linguagem de programação similar à linguagem S e ambiente S-Plus, que fornece uma ampla variedade de técnicas estatísticas (modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais etc), permite manipular dados e gráficos com grande facilidade etc. Além disso, existem alguns pacotes (*library*) que podem ser usados para certas análises estatísticas escritas apenas para R (ou S), Thompson (2008).

Uma cópia do *software* R pode ser obtido no *site* do CRAN: <<http://cran.r-project.org>>.

Quando se trata da análise de dados utilizando modelos lineares mistos, há diversas facilidades no *software* R, que apresenta um eficiente pacote **nlme** (acrônimo para modelos mistos não lineares), que apesar do nome, inclui facilidades de instalações para modelos lineares mistos através da função **lme()**.

Uma descrição mais detalhada das várias funções, classes e métodos disponíveis para uso dos pacotes do *software* R pode ser encontrado no seu arquivo *help*, tendo atualmente um total de 1853 pacotes disponíveis, o que inclui pacote **nlme**, com descrições. Existe também no *site* do projeto R, um sistema de busca com uma base ainda maior de informações sobre a linguagem.

A procura por ajuda em uma lista de discussão sobre o R pode ser feita na *R-help*-Lista de discussão internacional e R-STAT - lista nacional, com cadastro através da página do

grupo: Yahoo Grupos.

Um dos exemplos de dados de medidas repetidas, disponíveis pelo pacote `nlme` e exemplificado através do arquivo *help* com diferentes utilidades da função `lme()`.

Considerando o conjunto de dados `Orthodont` de um estudo presente em Potthoff e Roy (1964), que consiste de quatro medidas da distância em milímetros do centro da pituitária à fissura pteromaxilar feita aos 8, 10, 12 e 14 anos de idade de 16 garotos e 11 garotas.

```
> require(nlme)
> Orthodont
Grouped Data: distance~age|Subject
  distance age Subject  Sex
1      26.0   8      M01 Male
2      25.0  10      M01 Male
3      29.0  12      M01 Male
4      31.0  14      M01 Male
...
```

A saída pelo R apresenta a fórmula `distance~age|Subject` baseada nas colunas nomeadas:

- **distance:** a distância do centro da pituitária à fissura pteromaxilar em mm;
- **age:** a idade do indivíduo quando foi realizada a medida;
- **Subject:** um fator indicando em qual indivíduo a medida foi feita;
- **Sex:** um fator indicando se o indivíduo é gatoro ou garota, ou seja, male ou female, respectivamente.

Os modelos lineares mistos descritos por Laird e Ware (1982) são ajustados com a função `lme()`, usando o método da máxima verossimilhança - MV ou máxima verossimilhança restrita - MVR. Vários argumentos podem ser usados com esta função, sendo o mais típico:

```
>lme(fixed, data, random, correlation, weights, method)
```

O argumento:

- **fixed:** descreve parte do modelo relativa aos efeitos fixos, que devem ser declarados como objetos `groupedData` ou `lme.lmList`, que são implementados separadamente;
- **data:** indica o nome do arquivo de dados que contém as variáveis nomeadas como fixas, aleatórias, covariáveis etc;

- **random**: contém uma fórmula especificando os efeitos da parte aleatória do modelo;
- **correlation**: argumento opcional utilizado para descrever a estrutura de correlação intra-indivíduos. Uma lista de opções está disponível na classe **corStruct**. Por *default* admite-se correlações nulas intra-indivíduos;
- **weights**: argumento opcional utilizado para descrever a estrutura heterocedástica intra-indivíduos. Por *default* admite-se a homoscedasticidade dos erros intra-indivíduos.
- **method**: serve para especificar o método a ser utilizado na estimação do modelo linear misto. Especificando **method=REML**, o modelo é ajustado pelo método da máxima verossimilhança restrita. Se **method=ML**, o modelo é ajustado pelo método da máxima verossimilhança.

Há vários métodos disponíveis para o ajuste de objetos com a função **lme()**, incluindo aqueles desenvolvidos para funções genéricas como **anova()**, **print()**, **summary()** e **plot()**. Além disso, a função **lme()** inclui os comandos **fixed.effects** e **random.effects** usados para exibir as estimativas dos efeitos fixos e dos efeitos aleatórios, respectivamente.

2.2.1 Estruturas da Matriz Positiva Definida (pdMat)

Diferentes estruturas de matriz positiva definida podem ser usadas para representar a matriz de covariância **D**, de efeitos aleatórios com a função **lme()** que estão organizados em diferentes códigos na classe **pdMat**. A Tabela 1 lista as classes **pdMat** disponíveis para **lme()**. Por *default*, a classe **pdSymm** é usada para representar a matriz de efeitos aleatórios pelo argumento **random**, correspondendo a matriz não estruturada.

A seguir é apresentado um exemplo da matriz **D** associada a um modelo linear misto com dois efeitos aleatórios associados ao i -ésimo indivíduo.

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0t} \\ \sigma_{b_0t} & \sigma_t^2 \end{pmatrix}$$

Tabela 1 - Classes de estruturas das matrizes de covariância (pdMat) positivas definidas

Classe	Descrição
pdSymm	Positiva-definida geral
pdDiag	Diagonal
pdIdent	Multipla da identidade
pdCompSymm	Simetria composta
pdBlocked	Bloco diagonal

Fonte: Pinheiro e Bates (2000)

2.2.2 Estruturas de Correlação e Função de Variância

A matriz de covariância intra-indivíduos, $\mathbf{R}_i$, relacionada com o modelo (3), pode ser decomposta em um produto de matrizes mais simples:

$$\mathbf{R}_i = \mathbf{V}_i \mathbf{C}_i \mathbf{V}_i$$

onde $\mathbf{V}_i$ é uma matriz diagonal que descreve a variância dos erros intra-indivíduos e $\mathbf{C}_i$ é uma matriz de correlação positiva definida com todos os elementos da diagonal iguais a 1. A matriz $\mathbf{V}_i$ não é única e para assegurar unicidade, os elementos na diagonal devem ser positivos. Assim, verifica-se que

$$Var(\epsilon_{ij}) = \sigma^2 [\mathbf{V}_i]_{jj}^2, \quad cor(\epsilon_{ij}, \epsilon_{jk}) = [\mathbf{C}_i]_{jk}$$

Esta decomposição apresentada por Pinheiro e Bates (2000) é conveniente teorica e computacionalmente, permitindo desenvolver códigos ou classes da função `lme()` para as duas estruturas separadamente e combiná-las para obter uma família flexível de estruturas de variâncias e covariâncias.

Para modelar a estrutura de variância de covariâncias intra-indivíduos usando covariadas (variáveis independentes), Davidian e Giltinan (1995) apresentam a definição da variância dos erros intra-indivíduos:

$$var(\epsilon_{ij}|b_i) = \sigma^2 g(\mu_{ij}, v_{ij}, \delta), \quad i = 1, \dots, c, \quad j = 1, \dots, n_i$$

em que $\mu_{ij} = E(y_{ij}|b_{ij})$, v_{ij} é um vetor de covariadas da variância, δ é um vetor de parâmetros da variância e $g(\cdot)$ é uma função de variância. Por *default* os erros intra-indivíduos são assumidos independentes e homocedásticos, ou seja, $\mathbf{R}_i = \sigma^2 \mathbf{I}$ conhecida como Componente de Variância (VC).

Tabela 2 - Classes das funções de variância (**varFunc**)

Classe	Descrição
varExp	Exponencial da covariante da variância
varPower	Potência da covariante da variância
varConstPower	Constante somada a uma potência da covariante de variância
varIdent	Diferentes variâncias por níveis de um fator
varFixed	Pesos fixos, determinado por covariante de variância
varComb	Combinação de funções de variância

Fonte: Pinheiro e Bates (2000)

Tabela 3 - Classes das estruturas de correlação (**corStruct**)

Classe	Descrição
corAR1	AR(1)
corARMA	ARMA(p,q)
corCAR1	AR(1) contínua
corCompSymm	Simetria Composta
corExp	Exponencial - correlação espacial
corGauss	Gaussiana - correlação espacial
corLin	Linear- correlação espacial
corRation	Quadrática Racional-correlação espacial
corSpher	Esférica- correlação espacial
corSymm	Matriz de correlação geral

Fonte: Pinheiro e Bates (2000)

A estrutura da matriz de covariâncias intra-indivíduo, $\mathbf{R}_i$, pode ser flexivelmente modelada usando a função `lme()` e a combinação das estruturas de correlação, $\mathbf{C}_i$, com as funções de variância $\mathbf{V}_i$, que são organizadas nas classes **corStruct** e **varFunc**, respectivamente. Tabelas 2 e 3 listam as classes usuais para cada uma delas.

Utilizando os dados **Orthodont**, Pinheiro e Bates (1999) apresentam o ajuste de modelos com a função `lme()`, combinando classes de correlação e classes das funções de variância. Um exemplo pode ser dado por:

```
f<-lme(distance~age*Sex,data=Orthodont,
        random=pdDiag(~age),
        weights=varIdent(form=~1|Sex),
        correlation=corAR1())
```

O primeiro argumento é uma fórmula especificando o modelo, tendo interesse na diferença das restas associadas aos garotos e às garotas. Os dados são especificados pelo objeto `Orthodont` através do argumento `data`. Ao utilizar `pdDiag(~age)` o modelo admite interceptos independentes e efeito linear da idade como efeitos aleatórios. Os principais argumentos das funções `varFunc` e `corStruct` são `value` e `form` e são testadas as estruturas de variância `varIdent` combinadas com a estrutura de correlação `corAR1()`.

2.3 O Software SAS

Uma outra ferramenta computacional importante na análise de dados com medidas repetidas, utilizando modelos lineares mistos, é o `proc mixed` do SAS® (*Statistical Analysis System*), que possibilita a escolha de diversas estruturas para as matrizes de covariâncias $\mathbf{D}$ e $\mathbf{R}_i$. Este procedimento surgiu na década de 90 e permite especificar a estrutura da matriz $\mathbf{D}$, associada aos efeitos aleatórios através do comando `random`, e a matriz $\mathbf{R}_i$, associada aos efeitos fixos através do comando `repeated`.

Na análise de dados que apresentam medidas repetidas no tempo, a estrutura de covariâncias entre tempos é uma estrutura associada aos efeitos fixos, como pode ser visto em Verbeke e Molenberghs (1997) e Littell et al. (1998).

Grande parte dos trabalhos valoriza a capacidade do `proc mixed` do SAS na análise de dados utilizando-se modelos lineares mistos. Outros *softwares*, como o Stata e S-Plus, também trabalham muito bem esses modelos.

Há várias fontes para aperfeiçoar os conhecimentos sobre o uso do SAS, como livros, *sites* (<http://www.sas.com>, dentre outros), manuais e listas de discussão como SAS-L (<http://www.listserv.uga.edu/archives/sas-l.html>) e SAS Brasil (cadastro no *site* Yahoo Grupos-SASBrasil).

2.3.1 Estruturas da Matriz de Covariâncias

Algumas das estruturas mais utilizadas para as matrizes de covariâncias $\mathbf{D}$ e $\mathbf{R}_i$ e que já se encontram implementadas no `proc mixed` são apresentadas na Tabela 4. Uma lista completa pode ser encontrada no *site* do SAS ou no seu próprio *help*.

Tabela 4 - Algumas das estruturas de covariância disponíveis pelo `proc mixed` do SAS

Estrutura	Descrição	Número de Parâmetros	(i,j) -ésimo elemento
AR(1)	Autoregressiva(1)	2	$\sigma^2 \rho^{ i-j }$
ARH(1)	Heterogênea AR(1)	n_i+1	$\sigma_i \sigma_j \rho^{ i-j }$
CS	Simetria Composta	2	$\sigma_1 + \sigma^2 1(i=j)$
CSH	CS Heterogênea	n_i+1	$\sigma_i \sigma_j [\rho 1(i \neq j) + 1(i=j)]$
HF	Huynh-Feldt	n_i+1	$(\sigma_i^2 + \sigma_j^2)/2 + \lambda 1(i \neq j)$
UN	Não Estruturada	$n_i(n_i+1)/2$	σ_{ij}
VC	Componentes de Variância	1	$\sigma_k^2 1(i=j)$ i corresponde ao k -ésimo efeito

Nota: n_i -número de ocasiões.

Tendo em vista que o número de estruturas de covariâncias é elevado, um dos principais objetivos da análise com `proc mixed` do SAS é o de buscar, dentre várias estruturas possíveis, uma estrutura parcimoniosa que melhor represente os dados. No presente trabalho essas implementações serão feitas no `proc mixed` do SAS versão 9.1 para Windows.

3 MATERIAL E MÉTODOS

3.1 Material

O estudo dos procedimentos de investigação está apoiado na análise de dados da tese de Fernandes (2007) que avaliou os efeitos da Contagem de Células Somáticas - CCS nas frações de caseína do leite UAT (Ultra Alta Temperatura), ao longo do tempo de armazenamento. As variáveis utilizadas no presente trabalho foram as concentrações em (mg/mL) de α_{S1} -caseína e beta-caseína.

A CCS dos leites utilizados na fabricação dos 15 lotes de leite UAT variou de 197.000 a 800.000 CS/mL ou 5,29 a 5,90 log CS/mL. O grupo com baixa CCS (Grupo 1) constitui-se de lotes com contagens entre 197.000 a 316.000 CS/mL; o grupo com CCS intermediária (Grupo 2), por lotes com contagens entre 379.000 a 560.000 CS/mL e o grupo com alta CCS, (Grupo 3) constitui-se de lotes com contagem entre 600.000 a 800.000 CS/mL.

O delineamento experimental utilizado foi o de blocos casualizados, com cinco repetições. Foram utilizadas cinco medidas repetidas ao longo do período de armazenamento, aos 8, 30, 60, 90 e 120 dias.

As amostras de leite UAT foram coletadas em uma usina de beneficiamento localizada no Município de Casa Branca-SP, no período de março de 2005 a março de 2006.

As médias da concentração de α_{S1} -caseína e β -caseína em mg/mL dos lotes de leite UAT por dia e respectivos erros padrões, são apresentados na Tabela 5, assim como os gráficos de perfis individuais (Figuras 2 e 3, respectivamente) e perfis médio pela Figura 4.

Tabela 5 - Médias e erros padrões (e.p.) da concentração de α_{S1} -caseína e β -caseína em mg/mL do leite AUT nos cinco tempos de estocagem por grupo

α_{S1} – caseína						
Grupo 1		Grupo 2		Grupo 3		
Dia	Média	e.p.	Média	e.p.	Média	e.p.
8	11,30	2,18	10,23	2,60	8,70	4,28
30	8,94	4,52	9,76	3,33	9,32	3,36
60	9,21	2,07	9,07	2,83	8,37	2,22
90	7,10	6,02	7,50	4,45	7,16	3,82
120	6,99	3,62	6,95	4,14	6,65	3,94

β – caseína						
Grupo 1		Grupo 2		Grupo 3		
Dia	Média	e.p.	Média	e.p.	Média	e.p.
8	11,54	2,47	11,32	2,47	9,58	3,57
30	11,24	3,37	10,53	3,37	8,27	4,60
60	10,56	0,55	9,88	0,55	8,38	3,88
90	9,43	1,74	8,47	1,74	6,97	4,76
120	7,45	3,00	7,81	3,00	6,39	4,71

Figura 2 - Perfis individuais das concentração de α_{S1} -caseína (mg/mL) do leite UAT nas cinco ocasiões de armazenamento, por grupo

Figura 3 - Perfis individuais das concentrações de β -caseína (mg/mL) do leite UAT nas cinco ocasiões de armazenamento, por grupo

Figura 4 - Perfis médios de resposta da concentração de α_{S1} -caseína e β -caseína do leite UAT por grupo

Um terceiro conjunto de dados utilizados para ilustrar as técnicas de análise propostas são da dissertação de Lima (1988), que estudou o ajuste de um modelo linear multivariado de crescimento de aves. Foram utilizados os pesos de 32 frangos de corte da linhagem *Hubbard* (13 fêmeas e 19 machos), alojados em dois boxes, separados por sexo e alimentados com a mesma ração comercial. As aves foram identificadas por um anel de alumínio numerado colocado em sua asa direita. Cada ave foi pesada semanalmente durante um período de sete semanas, sendo as avaliações feitas sempre nos mesmos horários e dias da semana. Os pesos médios e os erros padrões das aves estão apresentados na Tabela 6.

Tabela 6 - Médias e erros padrões (e.p.) dos pesos em gramas, de frangos de corte *Hubbard*, por sexo e semana de idade

Semana	Fêmea		Macho	
	Média	e.p.	Média	e.p.
1	133,30	36,42	129,53	65,00
2	321,38	81,71	324,26	120,56
3	559,92	200,02	621,79	179,23
4	807,54	255,97	895,42	286,15
5	1089,85	193,47	1241,58	514,22
6	1473,00	271,49	1702,47	582,65
7	1770,00	298,41	2067,37	714,59

Figura 5 - Perfis individuais de peso corporal dos frangos de corte durante sete semanas de idade

Figura 6 - Perfil médio do peso corporal de frangos de corte *Hubbard*, por sexo e por idade

Na Tabela 6 como no gráfico da Figura 6 pode-se perceber um aumento na variabilidade dos pesos dos machos e fêmeas com o aumento da idade das aves.

3.2 Métodos

Uma abordagem simples de construção de um modelo linear misto a ser usado na análise de um conjunto de dados longitudinais consta das seguintes etapas: identificação dos efeitos aleatórios (associados aos indivíduos), escolha dos efeitos fixos (associados aos perfis médios de resposta) e a escolha da melhor estrutura de covariâncias entre e intra-indivíduos.

A sugestão de inclusão de efeitos aleatórios no modelo é feita a partir da análise do gráfico de perfis individuais de resposta para cada um dos grupos. Já o grau do polinômio a ser ajustado para explicar o comportamento das respostas médias ao longo do tempo, foi sugerido pelo gráfico de perfis médios de resposta. Sugestões de estruturas de covariâncias são obtidas a partir da análise dos gráficos de perfis individuais e da estimativa da matriz de covariâncias dos dados originais.

Tanto as inferências sobre os parâmetros de efeito aleatório e de efeito fixo, quanto para as comparações das diferentes estruturas de covariâncias para as matrizes $\mathbf{D}$ e $\mathbf{R}_i$, serão baseadas nos resultados do Teste da Razão de Verossimilhança e nos critérios de informação de Akaike e Bayesiano.

• **Concentração de α_{S1} -caseína no leite UAT (mg/mL)**

A partir da análise dos gráficos de perfis individuais (Figura 2) e médios Figura 4 (a), supõe-se que a concentração de α_{S1} -caseína no leite UAT do i -ésimo lote medida no tempo de estocagem t_{ij} pode ser bem explicada pelo modelo:

$$y_{ij} = \overbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 G1_i + \beta_3 G2_i + \beta_4 G1_i t_{ij} + \beta_5 G2_i t_{ij}}^{\text{efeitos fixos}} + \underbrace{b_{0i} + b_{1i} t_{ij} + \epsilon_{ij}}_{\text{efeitos aleatórios}} \quad (7)$$

para $i = 1, \dots, 15$, $j = 1, \dots, 5$ e $t_{ij}=8, 30, 60, 90$ e 120 , com a variável indicadora, Gl_i , assumindo os valores:

$$Gl_i = \begin{cases} 1 & \text{se } y_{ij} \in \text{ao Grupo } l, l=1,2 ; \\ 0 & \text{caso contrário.} \end{cases}$$

Resumindo:

$$y_{ij} = \begin{cases} \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij}, & \text{se } \in \text{grupo 3} \\ \beta_0 + \beta_2 + b_{0i} + (\beta_1 + \beta_4 + b_{1i})t_{ij} + \epsilon_{ij}, & \text{se } \in \text{grupo 1} \\ \beta_0 + \beta_3 + b_{0i} + (\beta_1 + \beta_5 + b_{1i})t_{ij} + \epsilon_{ij}, & \text{se } \in \text{grupo 2} \end{cases}$$

Os parâmetros β_0 ao β_5 representam os efeitos fixos associados ao intercepto, às variáveis indicadoras de grupo e aos termos de interação do modelo. Os termos b_{0i} e b_{1i} representam efeitos aleatórios associados ao intercepto e ao efeito linear de dia, para cada lote i de leite UAT. Assume-se que $\mathbf{b}_i = (b_{0i}, b_{1i})^\top \sim N(\mathbf{0}, \mathbf{D})$, onde

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 t} \\ \sigma_{b_0 t} & \sigma_t^2 \end{pmatrix}$$

É uma matriz positiva definida com três parâmetros, denominados componentes de variância e de covariância. O erro ϵ_{ij} associado à observação y_{ij} tem distribuição $\epsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$, onde

$$\mathbf{R}_i = \begin{pmatrix} Var(\epsilon_{i1}) & cov(\epsilon_{i1}, \epsilon_{i2}) & \cdots & cov(\epsilon_{i1}, \epsilon_{i5}) \\ cov(\epsilon_{i1}, \epsilon_{i2}) & Var(\epsilon_{i2}) & \cdots & cov(\epsilon_{i2}, \epsilon_{i5}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{i1}, \epsilon_{i5}) & cov(\epsilon_{i2}, \epsilon_{i5}) & \cdots & Var(\epsilon_{i5}) \end{pmatrix}$$

• **Concentração de β -caseína no leite UAT (mg/mL)**

De maneira similar aos da variável α_{S1} -caseína, realiza-se a análise dos gráficos de perfis individuais (Figura 3) e médios Figura 4 (b), supõe-se que a concentração de β -caseína no leite UAT do i -ésimo lote medida no tempo de estocagem t_{ij} pode ser bem explicada pelo modelo:

$$y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 G1_i + \beta_4 G2_i + \beta_5 G1_i t_{ij} + \beta_6 G2_i t_{ij} + \beta_7 G1_i t_{ij}^2 + \beta_8 G2_i t_{ij}^2}_{\text{efeitos fixos}} + \underbrace{b_{0i} + b_{1i} t_{ij} + b_{2i} t_{ij}^2 + \epsilon_{ij}}_{\text{efeitos aleatórios}} \quad (8)$$

para $i = 1, \dots, 15$, $j = 1, \dots, 5$ e $t_{ij}=8, 30, 60, 90$ e 120 , com a variável indicadora, $G1_i$, assumindo os valores:

$$G1_i = \begin{cases} 1 & \text{se } y_{ij} \in \text{ao Grupo } l, l=1,2 ; \\ 0 & \text{caso contrário.} \end{cases}$$

Resumindo:

$$y_{ij} = \begin{cases} (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + (\beta_2 + b_{2i})t_{ij}^2 + \epsilon_{ij}, & \text{se } \in \text{grupo 3} \\ (\beta_0 + \beta_3 + b_{0i}) + (\beta_1 + \beta_4 + b_{1i})t_{ij} + (\beta_2 + \beta_7 + b_{2i})t_{ij}^2 + \epsilon_{ij}, & \text{se } \in \text{grupo 1} \\ (\beta_0 + \beta_4 + b_{0i}) + (\beta_1 + \beta_5 + b_{1i})t_{ij} + (\beta_2 + \beta_8 + b_{2i})t_{ij}^2 + \epsilon_{ij}, & \text{se } \in \text{grupo 2} \end{cases}$$

Os parâmetros β_0 a β_8 representam os efeitos fixos associados ao intercepto, às variáveis indicadoras de grupo e aos termos de interação do modelo. Os termos b_{0i} , b_{1i} e b_{2i} representam efeitos aleatórios associados ao intercepto específico e efeito linear de dia e termo quadrático de dia para cada lote i de leite UAT. Assume-se que $\mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})^\top \sim N(\mathbf{0}, \mathbf{D})$, onde

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 t} & \sigma_{b_0 t^2} \\ \sigma_{b_0 t} & \sigma_t^2 & \sigma_{t t^2} \\ \sigma_{b_0 t^2} & \sigma_{t t^2} & \sigma_{t^2}^2 \end{pmatrix}$$

O erro ϵ_{ij} associado a observação y_{ij} tem distribuição normal $\epsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$, onde

$$\mathbf{R}_i = \begin{pmatrix} Var(\epsilon_{i1}) & cov(\epsilon_{i1}, \epsilon_{i2}) & \cdots & cov(\epsilon_{i1}, \epsilon_{i5}) \\ cov(\epsilon_{i1}, \epsilon_{i2}) & Var(\epsilon_{i2}) & \cdots & cov(\epsilon_{i2}, \epsilon_{i5}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{i1}, \epsilon_{i5}) & cov(\epsilon_{i2}, \epsilon_{i5}) & \cdots & Var(\epsilon_{i5}) \end{pmatrix}$$

• **Peso corporal de frangos de corte da linhagem *Hubbard***

Este modelo inicial inclui os efeitos fixos de tempo, tempo quadrado, sexo, sexo interação por tempo e sexo interação com tempo quadrado. Também inclui três efeitos aleatórios associados a cada indivíduo: intercepto, o efeito linear e efeito quadrático de tempo. Isto permite que cada indivíduo possa ter uma única trajetória parabólica, com coeficientes que variam aleatoriamente em torno dos efeitos fixos que definem a curva média de crescimento para cada sexo.

$$y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 M_i + \beta_4 M_i t_{ij} + \beta_5 M_i t_{ij}^2}_{\text{efeitos fixos}} + \underbrace{b_{0i} + b_{1i} t_{ij} + b_{2i} t_{ij}^2 + \epsilon_{ij}}_{\text{efeitos aleatórios}} \quad (9)$$

com M_i a variável indicadora:

$$M_i = \begin{cases} 0 & \text{se for Fêmea;} \\ 1 & \text{se for Macho} \end{cases}$$

$$y_{ij} = \begin{cases} \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + (\beta_2 + b_{2i})t_{ij}^2 + \epsilon_{ij}, & \text{se for Fêmea;} \\ (\beta_0 + \beta_3 + b_{0i}) + (\beta_1 + \beta_4 + b_{1i})t_{ij} + (\beta_2 + \beta_5 + b_{2i})t_{ij}^2 + \epsilon_{ij}, & \text{se for Macho} \end{cases}$$

Tem-se que $i = 1, \dots, 32$ e $j = 1, \dots, 7$, ($n_i = 7$ para todo i) e $\mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})^\top \sim N(\mathbf{0}, \mathbf{D})$, onde

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 t} & \sigma_{b_0 t^2} \\ \sigma_{b_0 t} & \sigma_t^2 & \sigma_{t t^2} \\ \sigma_{b_0 t^2} & \sigma_{t t^2} & \sigma_{t^2}^2 \end{pmatrix}$$

O erro ϵ_{ij} associado a observação y_{ij} tem distribuição normal $\epsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$, onde

$$\mathbf{R}_i = \begin{pmatrix} Var(\epsilon_{i1}) & cov(\epsilon_{i1}, \epsilon_{i2}) & \cdots & cov(\epsilon_{i1}, \epsilon_{i7}) \\ cov(\epsilon_{i1}, \epsilon_{i2}) & Var(\epsilon_{i2}) & \cdots & cov(\epsilon_{i2}, \epsilon_{i7}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{i1}, \epsilon_{i7}) & cov(\epsilon_{i2}, \epsilon_{i7}) & \cdots & Var(\epsilon_{i7}) \end{pmatrix}$$

Todos os modelos lineares mistos serão ajustados pelos métodos da máxima verossimilhança e máxima verossimilhança restrita, utilizando a função `lme()` do *software* R.

Para todas as variáveis respostas ajustou-se inicialmente um modelo com todos os efeitos fixos e aleatórios sugeridos (*modelo maximal*) e a partir dos resultados das comparações com

outros modelos encaixados (com um número menor de efeitos aleatórios), baseada nos resultados de Testes da Razão de Verossimilhança e dos valores dos Critérios de Informação de Akaike - AIC e de Schwarz - BIC, escolheu-se o melhor modelo conjunto de efeitos aleatórios.

Inicialmente, para todas as variáveis respostas, ajustou-se (método MVR) um modelo com todos os efeitos fixos e aleatórios sugeridos anteriormente (*modelo maximal*), admitindo-se uma estrutura UN (não estruturada) para a matriz de covariâncias entre indivíduos, $\mathbf{D}$, e uma estrutura VC (componente de variância) para a matriz de covariâncias intra-indivíduos, ou seja, $\mathbf{R}_i = \sigma^2 \mathbf{I}$.

A partir das comparações com outros modelos encaixados (com um número menor de efeitos aleatórios), baseadas nos resultados de Testes da Razão de Verossimilhança e dos valores dos Critérios de Informação de Akaike - AIC e de Schwarz - BIC, escolheu-se o melhor modelo conjunto de efeitos aleatórios.

Ajustando-se este último modelo pelo método de estimação MV, foram identificados os efeitos fixos significativos, através do teste de Wald e do teste-t para efeitos fixos, além das estatísticas indicadas anteriormente.

Escolhido também o melhor conjunto de efeitos fixos, procedeu-se à escolha da melhor estrutura de covariâncias intra-indivíduos, $\mathbf{R}_i$. Esta seleção foi feita utilizando-se o método de estimação MV ou MVR, os Testes da Razão de Verossimilhança e os valores de AIC e BIC. As estruturas de covariância avaliadas foram: UN, ARH(1), ARMA(1,1), AR(1) e CS que estão disponíveis no `proc mixed` do SAS e no R. Como os dados de α_{S1} -caseína e β -caseína são irregulares no tempo, as estruturas ARH(1), ARMA(1,1) e AR(1) não foram utilizadas, porque impõem que intervalos de tempo sejam igualmente espaçados.

Algumas comparações sobre os efeitos fixos foram feitas através de testes da Razão de Verossimilhança, envolvendo modelo completo versus modelo reduzido.

Os contrastes de interesse foram testados pelo teste de Wald, com o intuito de verificar se há mudanças nas interpretações para os efeitos fixos do modelo. Finalizando a seleção quando todas as variáveis não significativas foram eliminadas, ou seja, a fase de modelagem deve ser repetida até que se obtenha um modelo bem representativo da situação estudada.

Os comandos utilizados no `proc mixed` do SAS e no R para realizar as análises propostas estão listados nos Anexos D ao I. As variáveis utilizadas nas análises foram definidas da seguinte forma:

Concentração de α_{S1} -caseína e β -caseína no leite UAT:

- **lote**: único indentificador para cada lote de leite UAT;
- **Grupo2.f**: Lotes divididos em três categorias com CCS alta, intermediária e baixa;
- **Dia**: tempo de armazenamento (8, 30, 60, 90, 120) de leite UAT;
- **alphas1CN**: concentração de proteína no leite UAT (variável resposta) medida a cada dia;
- **betaCN**: concentração de proteína no leite UAT (variável resposta) medida a cada dia.

Peso corporal de frangos de corte:

- **indiv**: único identificador para cada frango;
- **sex**: gêneros Fêmea e Macho;
- **tempo**: semanas (1, 2, 3, 4, 5, 6 e 7) em que ocorreram as pesagens;
- **peso**: variável resposta do peso em gramas.

O diagnóstico dos modelos ajustados foi feito através de métodos gráficos. Uma boa qualidade de ajuste é alcançada quando os gráficos de resíduos padronizados versus valores observados retratarem bem os pressupostos do modelo. Os resíduos no processo de modelagem estatística têm uma relação importante com a qualidade do ajuste feito, constituindo uma das etapas no processo de escolha do modelo adequado. São úteis também para indicar a presença de pontos aberrantes, que poderão ser influentes ou não.

Com o uso do *software* R, o argumento `id` especificou-se o valor crítico, para decidir se o resíduo deve ser considerado aberrante ou não. Resíduos padronizados com valores superiores a $1 - id/2$ ($id = 0,05$) do quantil da distribuição normal padrão foram considerados aberrantes ao nível de significância de 5%. Os números dos indivíduos (lote ou indiv) correspondentes são indicados nos gráficos.

A suposição de normalidade dos resíduos foi analisada pela função `qqplot()` do *software* R com a construção de gráficos, permitindo verificar se o modelo está bem ajustado.

Os resultados numéricos de todos os modelos analisados foram comparados a partir das suas implementações no R e no `proc mixed` do SAS. De uma forma simplificada, evidenciou-se como esses *softwares* lidam com a análise de dados com medidas repetidas, quando se utilizam modelos lineares mistos, observando as vantagens e as restrições no uso dessas ferramentas.

4 RESULTADOS E DISCUSSÃO

4.1 Concentração de α_{S1} -caseína no leite UAT

A Tabela 7 apresenta algumas medidas descritivas da concentração de α_{S1} -caseína, que foram medidas em amostras de leite UAT durante o tempo de estocagem (de 8 a 120 dias).

Tabela 7 - Valores médios, mínimo e máximo de α_{S1} -caseína (mg/mL) do leite UAT nos cinco tempos de estocagem

	Estocagem (dias)					Geral
	8	30	60	90	120	
Média	10,06	9,34	8,88	7,25	6,86	8,48
Mínimo	5,55	5,51	7,19	3,76	3,51	
Máximo	12,27	12,17	10,82	9,34	9,07	

De acordo com os dados da Tabela 7, globalmente a concentração de α_{S1} -caseína variou de 3,51 mg/mL a 12,27 mg/mL, com uma média de 8,481 mg/mL. Os valores mínimos e máximos mudaram apenas ligeiramente em cada dia e a média diminuiu a cada dia. Percebe-se que tanto a variabilidade dos dados (máximos-mínimos), quanto as médias diminuem com o tempo de estocagem. Embora não evidente, na Figura 2 (seção 3.1) percebe-se o mesmo comportamento nos perfis individuais de reposta.

A matriz abaixo, que representa acima da diagonal principal, as estimativas das variâncias e covariâncias entre as medidas tomadas nos instantes 8, 30, 60, 90 e 120 e abaixo da diagonal as estimativas das correlações entre essas medidas, pode fornecer uma indicação de utilizar a técnica de modelos lineares mistos.

$$\begin{pmatrix} \mathbf{1,99} & \mathbf{1,06} & \mathbf{0,73} & \mathbf{1,77} & \mathbf{1,89} \\ 0,44 & \mathbf{2,86} & \mathbf{1,16} & \mathbf{2,33} & \mathbf{0,97} \\ 0,48 & 0,64 & \mathbf{1,15} & \mathbf{1,62} & \mathbf{0,98} \\ 0,57 & 0,63 & 0,69 & \mathbf{4,72} & \mathbf{2,82} \\ 0,77 & 0,33 & 0,52 & 0,74 & \mathbf{3,06} \end{pmatrix}$$

Pode-se perceber que as correlações entre as diferentes ocasiões medidas no tempo e diferentes variâncias ao longo do tempo, o que permite a proposta de diferentes estruturas através do estudo de modelos lineares mistos.

A tendência temporal dos dados da concentração de α_{S1} -caseína (em mg/mL) de cada Grupo (1=baixa, 2=intermediária e 3=alta CCS), pode ser avaliada na Figura 2, que mostra uma substancial variação de concentração de α_{S1} -caseína de lote para lote de leite UAT, dentro de cada grupo, além de uma variabilidade inicial nas respostas o que sugere considerar um efeito aleatório de intercepto para cada lote de leite UAT. Pode-se perceber um decréscimo consistente na concentração de α_{S1} -caseína nos lotes de leite UAT ao longo do tempo de armazenamento, sendo que em alguns lotes de leite esse decréscimo é mais acentuado. Pela Figura 4 (seção 3.1), os perfis médios mostram que a diminuição da concentração média de α_{S1} -caseína com o tempo de armazenamento, pode ser bem explicada por retas de inclinação possivelmente diferentes, que implica incluir a interação Dia e Grupo no modelo.

Iniciou-se o ajuste do modelo baseado na equação (7), que permite que cada lote possa ter uma reta, com coeficientes (intercepto e coeficiente angular) que variam aleatoriamente em torno do efeito fixo.

Os passos da análise dos dados de leite UAT estão representados na Tabela 8, no qual apresenta as estratégias de modelagem bem como cada modelo estipulado para análise. São apresentados a implementação de cada modelo pelo *software* R no anexo D e pelo *software* SAS no anexo E.

Tabela 8 - Relação de modelos lineares mistos utilizados na análise de dados de leite UAT em concentração de α_{S1} -caseína (mg/mL)

Modelo	Método Estimação	Efeito Fixo (*)	Efeito Aleatório	Estrutura R_i
A1	MVR	Dia, Grupo e interação	intercepto e Dia	VC
A2	MVR	Dia, Grupo e interação	Dia	VC
A3	MVR	Dia, Grupo e interação	intercepto	VC
A3.1	MV	Dia, Grupo e interação	intercepto	VC
A4	MV	Dia e Grupo	intercepto	VC
A4.1	MVR	Dia e Grupo	intercepto	VC
A5	MVR	Dia e Grupo	(-)	UN
A6	MVR	Dia e Grupo	intercepto	CS
A7	MV	Dia	intercepto	VC
A7.1	MVR	Dia	intercepto	VC

Nota: (*) intercepto.

(-) sem efeito aleatório.

MVR- Máxima Verossimilhança Restrita.

MV- Máxima Verossimilhança.

VC- componente de variância.

UN- não estruturada.

CS- simetria composta.

Os algoritmos de estimação dos parâmetros dos modelos por métodos de máxima

verossimilhança e máxima verossimilhança restrita no *software* R para a função `lme()`, são baseados em métodos computacionais iterativos de *Ridge-Stabilized Newton-Raphson* e *Fisher scoring* e sendo assim, as estimativas podem não convergir para uma solução (versão 2.8.1 do R). Após a implementação do modelo A1, a seguinte mensagem é exibida:

```
Erro em lme.formula(alphas1CN~Dia+Grupo2.f+Dia:Grupo2.f, random=~Dia,:
  nlminb problem, convergence error code=1
  message=iteration limit reached without convergence (9)
```

Como resultado, o objeto A1 não é criado, e as estimativas dos parâmetros no modelo A1 não podem ser obtidas. Uma alternativa foi utilizar o argumento `optim` da função `control=lmeControl(opt="optim")` baseado na otimização de Nelder-Mead, quasi-Newton e algoritmos de gradiente-conjugado, que é parte de uma lista de argumentos convenientes para o modelo e pode substituir o algoritmo padrão no R.

```
Linear mixed-effects model fit by REML (1)
Data: UATa2
      AIC      BIC    logLik (2)
 296.2789 318.62 -138.1395
Random effects: (3)
Formula: ~Dia | lote
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev      Corr (4)
(Intercept) 0.975446612 (Intr)
Dia          0.007064063 0.426
Residual     1.099880993
Fixed effects: alphas1CN ~ Dia + Grupo2.f + Dia:Grupo2.f (5)
              Value Std.Error DF   t-value p-value
(Intercept)   9.425058 0.5934272 57 15.882417 0.0000
Dia           -0.022491 0.0063153 57 -3.561394 0.0008
Grupo2.f1      1.495922 0.8392328 12  1.782488 0.1000
Grupo2.f2      1.203449 0.8392328 12  1.433988 0.1771
Dia:Grupo2.f1 -0.013544 0.0089312 57 -1.516511 0.1349
Dia:Grupo2.f2 -0.008783 0.0089312 57 -0.983442 0.3295
Correlation: (6)
              (Intr) Dia      Grp2.1 Grp2.2 D:G2.1
Dia           -0.335
Grupo2.f1     -0.707  0.237
Grupo2.f2     -0.707  0.237  0.500
Dia:Grupo2.f1  0.237 -0.707 -0.335 -0.167
Dia:Grupo2.f2  0.237 -0.707 -0.167 -0.335  0.500
Standardized Within-Group Residuals:
              Min      Q1      Med      Q3      Max
-3.2407592 -0.4491375  0.0649646  0.4183970  2.3423612
Number of Observations: 75
Number of Groups: 15
```

As estimativas dos parâmetros no modelo A1 são obtidas utilizando-se a função `summary()` e os resultados são apresentados no quadro, que mostra uma saída do modelo A1 contendo:

- (1) Indicação que o modelo foi ajustado por REML (método da máxima verossimilhança restrita) e que o conjunto de dados utilizados é o UATa2;
- (2) Várias medidas da qualidade do ajuste, incluindo AIC, BIC e logaritmo da verossimilhança;
- (3) Os efeitos aleatórios incluídos no modelo (intercepto e efeito linear de Dia) e as estruturas de covariâncias associadas;
- (4) A raiz quadrada das estimativas dos componentes de variância e covariância associados aos efeitos aleatórios e ao erro;
- (5) Apresenta ainda as estimativas dos parâmetros de efeito fixo, com os respectivos erros padrões e informações sobre o teste t para a hipótese de que cada parâmetro pode ser considerado nulo;
- (6) Estimativas das correlações entre parâmetros de efeito fixo e estatísticas descritivas dos resíduos intra-indivíduos.

Figura 7 - Retas ajustadas aos perfis individuais de α_{S1} -caseína de leite UAT-modelo A1

A Figura 7 é obtida com o uso da função **augPred** depois de ajustado o modelo A1, relacionado na Tabela 8. Nesta figura são visualizadas as retas ajustadas aos dados de cada um dos lotes. Essas retas apresentam interceptos e coeficientes angulares distintos, confirmando a necessidade de incluir o intercepto e o efeito linear de dia como aleatórios no modelo.

Na comparação do modelo A1 com o modelo A2 testa-se a inclusão do intercepto como efeito aleatório. Essa comparação feita pelo Teste da Razão de Verossimilhança - MVR resultou significativa (p-valor = 0,0112), o que implica em manter o intercepto como efeito aleatório. A hipótese nula e a hipótese alternativa envolvidas nessa comparação podem ser definidas em termos da matriz **D** como segue:

Hipótese:

$$H_0: \mathbf{D} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_t^2 \end{pmatrix}$$

$$H_a: \mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0t} \\ \sigma_{b_0t} & \sigma_t^2 \end{pmatrix}$$

Tabela 9 - Estatística de ajuste dos modelos A1 ao A4 (α_{S1} -caseína)

Modelos	Estimação	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
A1	MVR	10	296,2789	318,6200	276,279			
A2	MVR	8	301,2630	319,1358	285,263	A1×A2	8,98	0,0112
A3	MVR	8	293,8571	311,7299	277,857	A1×A3	1,58	0,4543
A3.1	MV	8	270,2920	288,8319	254,292			
A4	MV	6	269,3404	283,2454	257,3404	A3.1×A4	3,05	0,2178

A manutenção do efeito linear de Dia como aleatório é testada comparando-se os modelos A1 e A3. Como o teste resultou não significativo (p-valor=0,4543), pode-se omitir este efeito aleatório de todos os modelos posteriores.

Tabela 10 - Testes de Wald para os parâmetros de efeitos fixos do modelo A3.1

	gl num	gl den	F-valor	p-valor
(Intercepto)	1	57	696,86	<, 0000
Dia	1	57	81,91	<, 0001
Grupo2.f	2	12	0,47	0,6350
Dia:Grupo2.f	2	57	1,43	0,2458

Como os parâmetros de efeitos fixos Grupo2.f e Dia:Grupo2.f foram considerados nulos pelo teste de Wald (p-valor=0,6350 e 0,2458, respectivamente) criou-se o modelo A4 que não inclui estes termos.

A comparação dos modelos A3.1 e A4 pelo Teste da Razão de Verossimilhança (MV) também indica que o efeito fixo de Grupo2.f e o efeito da interação Dia:Grupo2.f não precisam ser mantidos no modelo.

Em seguida, ajusta-se o modelo A4 por máxima verossimilhança restrita obtendo-se o modelo denominado A4.1, que será usado em outras comparações com modelos que usam outras estruturas de covariâncias.

4.1.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$)

Escolhidos os efeitos fixos e o efeito aleatório (modelo A4), foram testadas as estruturas UN e CS para a matriz de covariância $\mathbf{R}_i$. Os ajustes dos modelos A5 e A6 também foram implementados no `proc mixed` do SAS. Para o ajuste do modelo A5 (estrutura UN) utilizou-se a função `gls()` do R. A Tabela 11 apresenta os valores dos critérios AIC e BIC, o valor $-2\log\text{Veross}$ para cada estrutura.

Tabela 11 - Estatísticas de ajuste dos modelos A4.1, A5 e A6 (α_{S1} -caseína)

Modelo	$\mathbf{R}_i$	gl	AIC	BIC	$-2\log\text{Veross}$	Comparação	χ^2	p-valor
A4.1	VC	6	276,8811	290,4572	264,8812			
A5	UN	19	282,5215	325,5124	244,5215	A4.1 $\times$ A5	20,36	0,0866
A6	CS	7	278,8811	294,7199	264,8812	A5 $\times$ A6	20,36	0,0606

Baseado nos resultados dos testes da razão de verossimilhança e nos valores dos critérios AIC e BIC, a estrutura de covariância escolhida como adequada foi a VC, que tem um único parâmetro.

Tabela 12 - Níveis descritivos (p-valores) dos testes para efeitos fixos, admitindo diferentes estruturas de covariâncias para matriz $\mathbf{R}_i$

Causas de Variação	Estruturas		
	VC	UN	CS
Intercepto	<0,0001	<,0001	<,0001
Dia	<0,0001	<,0001	<,0001
Grupo2.f	0,6725	0,0641	0,6725

Na Tabela 12 pode-se perceber que foram encontrados p-valores diferentes para Grupo2.f entre os modelos propostos, mas as conclusões mantêm-se as mesmas, já que todos os p-valores foram superiores a 5%. Como o efeito de Grupo2.f resultou não significativo (p-valor > 0,05) e o efeito linear do tempo de armazenamento resultou significativo (p-valor < 0,001), construiu-se o modelo A7 que inclui somente o efeito linear do tempo na sua parte fixa.

Tabela 13 - Estatísticas de ajuste dos modelos A4 e A7 (MV) (α_{S1} -caseína)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
A4	6	269,3404	283,2454	257,3404			
A7	4	266,3325	275,6024	258,3324	A4×A7	0,99	0,6089

Da Tabela 13 percebe-se que os menores valores de AIC e BIC, além da não significância da comparação dos modelos A4 e A7 pelo Teste da Razão de Verossimilhança, indicam que o modelo A7 é o que melhor representa os dados de “ α_{S1} -caseína”.

As estimativas dos parâmetros da reta (modelo A7.1) que melhor representa a relação entre a α_{S1} -caseína e o tempo de armazenamento de leite UAT são apresentados na Tabela 14. Pode-se concluir que a concentração de α_{S1} -caseína (mg/mL) descreve a uma taxa constante de 0,0299 mg/mL por dia de armazenamento.

Tabela 14 - Estimativas dos parâmetros fixos do modelo A7.1

Parâmetros	Valor	Erro Padrão	gl	t-valor	p-valor
Intercepto	10,3248	0,3867	59	26,70	0,00
Dia	-0,0299	0,0033	59	-9,12	0,00

Com as estimativas dos parâmetros do modelo A7.1, pode-se obter a equação (Figura 8) que descreve a concentração de α_{S1} -caseína (mg/mL) com $t_{ij} \in [8, 120]$:

$$y_{ij} = 10,3248 - 0,0299t_{ij}$$

Segundo Fernandes (2004), a diminuição da concentração de α_{S1} -caseína deve-se a ação de proteases de origem dos leucócitos e das plasminas, uma vez que estas hidrolisam a fração citada. As porcentagens de α_{S1} -caseína no leite UAT não são afetadas pelo nível de células somáticas. Para maiores detalhes consulte Grieve e Kitchen, 1985 apud Fernandes (2004).

Figura 8 - Retas ajustadas para concentração de α_{S1} -caseína (mg/mL) no leite UAT durante o tempo de estocagem, para cada grupo-modelo A7.1

4.1.2 Diagnóstico do modelo A7.1

A verificação dos pressupostos do modelo A7.1 (modelo A7 ajustado por MVR) pode ser realizada através da análise gráfica dos resíduos com procedimentos do *software* R, que permitem afirmar se o modelo apresentou um bom ajuste. A Figura 9 (a) mostra que os resíduos padronizados do ajuste do modelo A7.1 se distribuem aleatoriamente em torno do valor zero. Percebe-se ainda a presença de alguns candidatos a *outliers*, ou seja, pontos com resíduos superiores, em valor absoluto a 2.

O q-q plot apresentado na Figura 9 (b) indica que a pressuposição de normalidade para erros intra-indivíduos é plausível. No entanto, a presença de alguns *outliers* pode justificar uma investigação mais aprofundada.

Finalmente, pode-se verificar uma boa concordância entre os valores observados de α_{S1} -caseína e os ajustados, com exceção de alguns *outliers* [Figura 9 (c)].

A distribuição das estimativas dos efeitos aleatórios (EBLUP's) gerada com o ajuste do modelo A7.1 pode ser avaliado no q-q plot apresentado na para Figura 9 (d). Nota-se que o

intercepto aleatório correspondente a uma observação com CCS= 630, pode ser um *outlier*. O julgamento da normalidade nos efeitos aleatórios fica prejudicado pela pouca quantidade deles.

Figura 9 - Diagnóstico do ajuste do modelo A7.1

4.2 Concentração de β -caseína no leite UAT

Tabela 15 - Valores médios, mínimo e máximo de β -caseína (mg/mL) do leite UAT nos cinco tempos de estocagem

	Estocagem (dias)					Geral
	8	30	60	90	120	
Média	10,81	10,01	9,60	8,29	7,21	9,18
Mínimo	7,04	5,57	5,88	3,97	3,64	
Máximo	12,81	12,22	10,97	10,53	8,97	

De acordo com os dados da Tabela 15, globalmente a concentração de β -caseína variou de 3,64 mg/mL a 10,815 mg/mL, com uma média de 9,188 mg/mL. Os valores mínimos e máximos mudaram apenas ligeiramente em cada tempo de estocagem e a média diminuiu ao longo do tempo. Percebe-se que tanto a variabilidade dos dados (máximos-mínimos), quanto as médias diminuem com o tempo de estocagem.

Nos gráficos com os perfis individuais das concentrações β -caseína avaliadas no intervalo de 8 a 120 dias de armazenamento (Figura 2 seção 3.1) percebe-se um decréscimo linear na resposta média. Sugerem ainda a utilização do intercepto e do efeito linear do tempo de estocagem para cada lote como efeitos aleatórios.

A matriz a seguir, apresenta acima da sua diagonal principal as variâncias das concentrações de β -caseína nos cinco tempos de estocagem; acima da diagonal são apresentadas as covariâncias e abaixo, as correlações entre essas medidas. Nela pode-se perceber uma pequena heterocedasticidade nas variâncias, algumas correlações altas e outras bem baixas, que sugerem que o seu estudo seja feito utilizando modelos lineares mistos.

$$\begin{pmatrix} \mathbf{1,33} & \mathbf{0,15} & \mathbf{0,29} & \mathbf{0,62} & \mathbf{0,56} \\ 0,08 & \mathbf{2,63} & \mathbf{0,75} & \mathbf{0,51} & \mathbf{0,30} \\ 0,18 & 0,34 & \mathbf{1,80} & \mathbf{1,98} & \mathbf{2,01} \\ 0,33 & 0,19 & 0,90 & \mathbf{2,65} & \mathbf{2,45} \\ 0,29 & 0,11 & 0,88 & 0,89 & \mathbf{2,87} \end{pmatrix}$$

Devido ao fato de os dados serem de medidas repetidas para cada unidade amostral, a pressuposição básica de independência nos modelos lineares deve ter sido violada.

Para uma visualização prática da situação, no gráfico da Figura 10 apresenta a

característica mais importante, que os resíduos correspondentes ao mesmo indivíduo tendem a ter o mesmo sinal (PINHEIRO; BATES, 2000; CALEGÁRIO, 2004). Esta característica motiva a usar modelagem de efeito misto, no caso em questão.

Figura 10 - Distribuição dos resíduos do modelo linear por lote de leite UAT

Figura 11 - Intervalos de confiança para intercepto e termo linear da reta para cada lote com $\gamma=95\%$

A Figura 11 foi construída a partir do ajuste de uma reta para cada lote e do cálculo de um intervalo de confiança para o intercepto e para a inclinação de cada reta. A

aparente variabilidade nesses intervalos indica a inclusão do intercepto e do efeito linear de tempo de estocagem como efeitos aleatórios no modelo linear misto. Com base nas análises anteriores, têm-se suficientes argumentos para usar modelos lineares mistos.

Tabela 16 - Relação dos modelos lineares mistos utilizados na análise de dados de concentração de β -caseína (mg/mL) no leite UAT

Modelo	Método Estimação	Efeito Fixo (*)	Efeito Aleatório	Estrutura $\mathbf{R}_i$
B1	MVR	Dia, Dia ² , Grupo e interações	intercepto, Dia e Dia ²	VC
B2	MVR	Dia, Dia ² , Grupo e interações	Dia e Dia ²	VC
B3	MVR	Dia, Dia ² , Grupo e interações	Dia	VC
B3.1	MV	Dia, Dia ² , Grupo e interações	Dia	VC
B4	MV	Dia, Dia ² , Grupo e Dia:Grupo	Dia	VC
B5	MV	Dia, Dia ² e Grupo	Dia	VC
B6	MV	Dia e Grupo	Dia	VC
B6.1	MVR	Dia e Grupo	Dia	VC
B7	MVR	Dia e Grupo	(-)	UN
B7.1	MV	Dia e Grupo	(-)	UN
B8	MVR	Dia e Grupo	Dia	CS
B9	MV	Grupo 1=2,3	(-)	UN
B9.1	MVR	Grupo 1=2,3	(-)	UN
B10	MV	Grupo 1=2=3	(-)	UN

Nota: (*) intercepto.

(-) sem efeito aleatório.

MVR- Máxima Verossimilhança Restrita.

MV- Máxima Verossimilhança.

VC- componente de variância.

UN- não estruturada.

CS- simetria composta.

As formas gerais dos modelos B1 ao B10 são mostradas na Tabela 16. Estas especificações correspondem à sintaxe utilizada para ajustar modelos lineares mistos com o `proc mixed` no SAS e a função `lme()` do R.

A partir do modelo B1 é retirado o efeito aleatório do intercepto, considerando que não existe grande variabilidade entre os lotes de leite no oitavo dia de armazenamento, o que é confirmado na Tabela 17 pelos menores valores dos critérios AIC e BIC associados ao modelo B2 e pelo alto p-valor da comparação dos modelos B1 e B2. Comparando os modelos B2 e B3 verifica-se que não há necessidade de manter no modelo o efeito quadrático de dia como aleatório, por conta dos menores valores dos critérios AIC e BIC associados ao modelo B3, e pelo alto p-valor da comparação dos modelos B2 e B3.

Tabela 17 - Estatísticas de ajuste dos modelos B1 ao B3 (MVR) (β -caseína)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
B1	16	335,5160	370,5504	303,5160			
B2	13	330,8740	359,3395	304,8740	B1 $\times$ B2	1,36	0,7154
B3	11	329,0544	353,1406	307,0544	B2 $\times$ B3	2,18	0,3362

Tabela 18 - Estatísticas de teste para os parâmetros de efeitos fixos dos modelos B3.1 ao B5

Modelo	variáveis	gl num	gl den	F-valor	p-valor
B3.1	Intercepto	1	54	3132,0291	<, 000
	Dia	1	54	45,9739	<, 0001
	I(Dia) ²	1	54	1,0879	0,3016
	Grupo2.f	2	12	15,9615	0,0004
	Dia:Grupo2.f	2	54	0,2967	0,7445
	I(Dia) ² :Grupo2.f	2	54	1,2849	0,2850
	Intercepto	1	56	3103,8535	<, 000
B4					
	Dia	1	56	46,7129	<, 0001
	I(Dia) ²	1	56	1,0688	0,3057
	Grupo2.f	2	12	15,8094	0,0004
	Dia:Grupo2.f	2	56	0,3015	0,7409
B5	Intercepto	1	58	3168,748	<, 000
	Dia	1	58	46,997	<, 0001
	I(Dia) ²	1	58	1,097	0,2993
	Grupo2.f	2	12	16,145	0,0004

Tendo decidido manter somente o efeito linear de Dia como aleatório, buscar-se-á estudar a parte fixa do modelo associado à concentração de β -caseína.

Para os testes dos parâmetros de efeito fixo do modelo, foram considerados os procedimentos descritos na seção 3.2. Os resultados correspondentes são apresentados nas Tabelas 18 e 19.

As estatísticas F apresentadas na Tabela 18, usadas para testar a significância da interação entre Dia² e Grupo no modelo B3.1, entre Dia e Grupo no Modelo B4 e Dia² no modelo B5, resultaram não significativas (p-valor = 0,2850, 0,7409 e 0,2993, respectivamente) indicando que sejam retiradas do modelo B3.

Os resultados de outra abordagem, feita através do Teste da Razão de Verossimilhança e dos critérios AIC e BIC, indicam o modelo B6 como o mais indicado para descrever a parte fixa do modelo associado à concentração de β -caseína.

Tabela 19 - Estatísticas de ajuste dos modelos B3.1 ao B6 (β -caseína)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
B3.1	11	256,8893	282,3817	234,8893			
B4	9	255,7407	276,5981	237,7407	B3.1 $\times$ B4	2,85	0,2403
B5	7	252,3982	268,6206	238,3982	B4 $\times$ B5	0,65	0,7198
B6	6	251,5619	265,4668	239,5619	B5 $\times$ B6	1,16	0,2807

Figura 12 - Relação entre os dados originais, o modelo B1 (pontilhado) e modelo B6.1 (linha contínua) para concentração de β -caseína (mg/mL).

4.2.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$)

O propósito final desta seção foi escolher a estrutura de covariâncias intra-indivíduos ($\mathbf{R}_i$) mais adequada, admitindo-se o modelo B6 para os efeitos fixos e aleatórios. Foram testadas somente as estruturas UN, associada ao modelo B7 e a estrutura CS (simetria composta), associada ao modelo B8.

Tabela 20 - Estatísticas de ajuste dos modelos B6.1 ao B8 (MVR) (β -caseína)

Modelos	$\mathbf{R}_i$	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
B6.1	VC	6	261,9551	275,5312	249,955			
B7	UN	19	258,5863	301,5772	220,5863	B6.1 $\times$ B7	29,36	0,0058
B8	CS	7	260,9537	276,7925	246,9537	B7 $\times$ B8	26,36	0,0095

Avaliando os resultados apresentados na Tabela 20, pode-se concluir que a estrutura UN (modelo B7) é a que melhor explica as variâncias e covariâncias intra-indivíduos, que apesar de não ter o menor valor de BIC, tem o menor valor de AIC e apresenta resultados favoráveis nas comparações realizadas pelo Teste da Razão de Verossimilhança.

Tabela 21 - Níveis descritivos (p-valores) dos testes para efeitos fixos, admitindo diferentes estruturas de covariâncias para a matriz $\mathbf{R}_i$

Causas de Variação	Estruturas		
	VC	UN	CS
Intercepto	<0,0001	<,0001	<,0001
Dia	<0,0001	<,0001	<,0001
Grupo2.f	0,0004	0,0003	0,0038

As diferentes estruturas de covariâncias intra-indivíduos devem ter alterados os valores dos erros padrões associados aos efeitos fixos, mas não alteraram os resultados das inferências (Tabela 21) realizados pelos testes F.

Tabela 22 - Estimativas dos parâmetros de efeito fixo do modelo B7

Parâmetro	Valor	Erro Padrão	t-valor	p-valor
Intercepto	10,1767	0,3800	26,7771	0,0000
Dia	-0,0395	0,0027	-14,6091	0,0000
Grupo2.f1	2,1485	0,5367	4,0025	0,0002
Grupo2.f2	1,7136	0,5368	3,1923	0,0021

As equações das restas ajustadas para a concentração de β -caseína (mg/mL) no leite UAT, assumindo o modelo B7 $t_{ij} \in [8, 120]$, podem ser escritas como:

$$y_{ij} = \begin{cases} 12,3253 - 0,039541t_{ij} & \text{Grupo}=1; \\ 11,8903 - 0,039541t_{ij} & \text{Grupo}=2; \\ 10,1767 - 0,039541t & \text{Grupo}=3. \end{cases}$$

Com base nestas equações e no gráfico da Figura 13 pode-se observar que com o decorrer do tempo de armazenamento, tem-se uma consequente diminuição linear nas concentrações de β -caseína.

Figura 13 - Retas ajustadas para a concentração de β -caseína (mg/mL) no leite UAT durante o tempo de estocagem, para cada grupo

Pela Figura 13, uma sugestão é realizar um teste afim de verificar se as retas associadas ao grupo 1 e 2 são coincidentes.

Com as informações apresentadas na Tabela 23, tem-se que a coincidência das retas associadas aos grupos 1 e 2 foi confirmada tanto pelo Teste da Razão de Verossimilhança (p-valor = 0,4888), que comparou os modelos B7.1 e B9, quanto pelos valores de AIC e BIC que foram favoráveis ao modelo B9. A comparação do modelo B10, que supõe uma única reta para os três grupos, com o modelo B9 resultou significativa (p-valor = 0,0010), indicando que o modelo B9 é o mais adequado para explicar o comportamento da concentração de β -caseína em função do tempo de estocagem. Os valores de AIC e BIC também foram favoráveis ao modelo B9.

Tabela 23 - Estatísticas de ajuste dos modelos B7.1 ao B10 (β -caseína)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
B7.1	19	247,9193	291,9516	209,9194			
B9	18	246,3986	288,1133	210,3986	B7×B9	0,4792	0,4888
B10	17	255,2785	294,6758	221,2786	B9×B10	10,8799	0,0010

A partir das estimativas dos parâmetros do modelo B9.1 (modelo B9 ajustado por MVR), obtém as equações das retas ajustadas, com $t_{ij} \in [8, 120]$:

$$y_{ij} = \begin{cases} 12,0710 - 0,0395t_{ij} & \text{Grupo=1 ou 2;} \\ 10,2257 - 0,0395t_{ij} & \text{Grupo=3.} \end{cases}$$

A Figura 14 mostra as duas retas ajustadas.

Figura 14 - Evolução da concentração de β -caseína (mg/mL) do leite UAT durante o armazenamento para cada grupo, segundo o modelo B9.1

Contudo o fator grupo atuando no intervalo de tempo tem impacto na qualidade do leite UAT e segundo Fernandes (2007), a CCS no leite cru utilizado para fabricação do leite UAT pode prejudicar a qualidade do produto, recomendando o uso do leite cru com baixa CCS (Grupo 3), afim de evitar possíveis efeitos da proteólise, obtendo-se um produto com qualidade adequada e estável ao longo do período de armazenamento.

4.2.2 Diagnóstico do modelo B9.1

Nesta seção é realizada uma avaliação gráfica do diagnóstico do modelo B9.1 (modelo B9 ajustado por MVR) utilizando o *software* R.

Figura 15 - Diagnóstico do ajuste do modelo B9.1 (β -caseína)

A Figura 15 (a) indica que o modelo B9.1 representa adequadamente a heterocedasticidade das respostas intra-lotes, por conta da distribuição aleatória dos resíduos ao redor do valor zero e da presença de somente dois valores discrepantes. A pressuposição de normalidade dos erros é aceitável diante da avaliação da Figura 15 (b).

4.3 Pesos de frango de corte da linhagem *Hubbard*

Os dados de peso dos frangos de corte da linhagem *Hubbard* avaliados semanalmente, até a 7^a semana de idade são balanceados em relação ao tempo, completos e regulares, favorecendo o ajuste de outras estruturas para a matriz de covariâncias intra-indivíduos. Os perfis individuais de resposta de 13 fêmeas (perfis 1 a 13) e 19 machos (perfis 14 a 32) sugerem a inclusão de três efeitos aleatórios no modelo misto: o intercepto, o efeito linear e o efeito quadrático do tempo.

Nas Figuras 5 (seção 3.1) e 16 percebe-se que os pesos dos machos são similares aos das fêmeas nas duas primeiras semanas, mas a diferença vai acentuando-se a partir da terceira semana. Nos boxplots da Figura 16 percebe-se que as variâncias dos pesos são crescentes ao longo das semanas.

Figura 16 - Boxplot dos pesos corporais dos frangos de corte, por sexo

A matriz apresentada a seguir mostra na diagonal principal as estimativas das variâncias dos pesos tomados nas sete ocasiões; acima diagonal, as covariâncias entre os pesos tomados nessas ocasiões e, abaixo da diagonal, as correlações entre essas medidas. Nela pode-se perceber um aumento acentuado da variância dos pesos ao longo do tempo e uma diminuição da correlação quando se aumenta o intervalo de semanas. Esses argumentos são suficientes para não se utilizar um modelo de regressão polinomial usual, que admite uma matriz de covariância $\sigma^2 \mathbf{I}$.

435,62	729,23	943,45	1248,72	1365,23	1147,27	1499,69
0,85	1661,23	2353,61	2897,68	2827,15	2746,97	2999,43
0,60	0,77	5613,84	6531,27	6030,51	6512,32	6123,29
0,56	0,66	0,81	11505,15	12067,12	12586,77	12045,92
0,42	0,45	0,52	0,73	23754,86	24996,73	27660,75
0,30	0,37	0,48	0,65	0,90	32471,89	36685,30
0,33	0,34	0,38	0,52	0,83	0,94	47164,03

Estes resultados indicam que o modelo de regressão linear não é adequado para representação do comportamento dos dados.

A Tabela 24 apresenta os diversos modelos lineares mistos a serem utilizados na análise dos dados de peso corporal dos frangos de corte. O modelo F1 inclui os efeitos fixos de

tempo, tempo², sexo e as interações entre sexo e tempo e entre sexo e tempo², isto é, sugere o ajuste de uma parábola para as respostas médias de peso ao longo do tempo, cada um dos sexos. Também inclui o intercepto e os efeitos linear e quadrático de tempo como efeitos aleatórios.

Tabela 24 - Relação dos modelos lineares mistos utilizados na análise dos dados de peso corporal dos frangos de corte

Modelo	Método Estimação	Efeito Fixo (*)	Efeito Aleatório	Estrutura R_i
F1	MVR	tempo tempo ² , sexo e interações	intercepto tempo tempo ²	VC
F2	MVR	tempo tempo ² , sexo e interações	tempo tempo ²	VC
F2.1	MV	tempo, tempo ² , sexo e interações	tempo tempo ²	VC
F3	MVR	tempo, tempo ² , sexo e interações	intercepto tempo	VC
F4	MV	tempo, tempo ² , sexo e tempo ² :sexo	tempo tempo ²	VC
F5	MV	tempo, tempo ² e tempo ² :sexo	tempo tempo ²	VC
F6	MV	tempo, tempo ² e tempo ² :sexo	(-)	UN
F6.1	MVR	tempo, tempo ² e tempo ² :sexo	(-)	UN
F7	MV	tempo, tempo ² e tempo ² :sexo	tempo tempo ²	ARH(1)
F8	MV	tempo, tempo ² e tempo ² :sexo	tempo tempo ²	AR(1)
F9	MV	tempo, tempo ² e tempo ² :sexo	tempo tempo ²	CS

Nota: (*) intercepto.

(-) sem efeito aleatório.

MVR- Máxima Verossimilhança Restrita.

MV- Máxima Verossimilhança.

VC- componente de variância.

UN- não estruturada.

ARH(1)-auto regressiva com heterogeneidade de variância.

AR(1)- auto regressiva.

CS- simetria composta.

Avaliando-se os boxplots dos resíduos por indivíduo apresentados na Figura 17, percebe-se que eles estão praticamente centrados no valor zero, mas que a variabilidade muda com o indivíduo. Para modelar esta heterocedasticidade do erro intra-indivíduo será utilizado o argumento `weights` na função `lme()`, que possibilita a especificação de diferentes estruturas para a matriz de covariâncias intra-indivíduo.

Figura 17 - Distribuição dos resíduos do modelo F1 por indivíduo

Para os testes de ajuste de modelos considerando alterações na parte aleatória, foram propostos três modelos através dos procedimentos descritos na seção 3.2, os resultados se encontram na Tabela 25.

Tabela 25 - Estatísticas de ajuste dos modelos de F1 ao F3 (MVR) (peso corporal de frangos)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
F1	13	2394,097	2438,096	2368,098			
F2	10	2394,091	2427,936	2374,092	F1×F2	5,99	0,1119
F3	10	2415,169	2449,014	2395,168	F1×F3	27,07	<,0001

Os valores dos critérios de informação AIC= 2394,091 e BIC=2427,936, referente ao modelo F2, sugerem ser o mais adequado para representar os dados no qual os efeitos de tempo e tempo² como aleatório foram mantidos, excluindo-se intercepto como aleatório, uma vez que os pesos iniciais dos frangos são muito parecidos e assim em todos os modelos posteriores.

Embora os resultados do teste de Wald apresentados na Tabela 26 tenham indicado um efeito significativo para a interação sexo e tempo (p-valor < 0,0001), as estatísticas de ajuste apresentadas na Tabela 27 sugerem a sua exclusão do modelo, pois os critérios AIC e BIC e o p-valor associado ao teste da razão de verossimilhança (comparação F2.1×F4) foram favoráveis ao modelo F4. Seguindo orientação de Pinheiro e Bates (2000) e de Verbeke e Molenberghs (2000), optou-se por utilizar os resultados no teste de razão de verossimilhança, excluindo a referida interação do modelo F2.1.

Tabela 26 - Estatísticas dos parâmetros de efeitos fixos dos modelos F2.1, F4 e F5

Modelo	variáveis	gl num	gl den	F-valor	p-valor
F2.1	Intercepto	1	188	1869,81	<, 0001
	tempo	1	188	5465,63	<, 0001
	I(tempo)^2	1	188	374,71	<, 0001
	sex	1	30	0,23	0,6368
	tempo:sex	1	188	29,01	<, 0001
	I(tempo)^2:sex	1	188	7,79	0,0058
F4	Intercepto	1	189	1876,72	<, 0001
	tempo	1	189	5490,36	<, 0001
	I(tempo)^2	1	189	376,00	<, 0001
	sex	1	30	0,23	0,6347
	I(tempo)^2:sex	1	189	36,76	<, 0001
F5	Intercepto	1	189	1880,792	<, 000
	tempo	1	189	5513,201	<, 0001
	I(tempo)^2	1	189	375,946	<, 0001
	I(tempo)^2:sex	1	189	36,142	<, 0001

O teste de Wald (Tabela 26) para o fator sexo resultou não significativo (p-valor=0,6347), concordando com os critérios AIC e BIC e o p-valor associado ao teste da razão de verossimilhança (comparação F4×F5) foram favoráveis ao modelo F5.

Tabela 27 - Estatísticas de ajuste dos modelos de F2.1, F4 e F5 (MVR) (peso corporal de frangos)

Modelos	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
F2.1	10	2421,535	2455,652	2401,53			
F4	9	2419,751	2450,456	2401,75	F2.1×F4	0,21	0,6425
F5	8	2418,790	2446,083	2402,79	F4×F5	1,04	0,3081

Definidos os efeitos que compõem a parte fixa e a parte aleatória do modelo misto, o estudo continua com a busca de uma estrutura adequada para a matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$). Nos testes da razão de verossimilhança apresentados na Tabela 28, em que se comparam os modelos F5 ao F9, utilizaram-se os resultados de ajustes por MV, e não MVR, pelo fato de ocorrer um erro na implementação desta opção na função `lme()` do R.

4.3.1 Ajuste de estruturas à matriz de covariâncias intra-indivíduos ($\mathbf{R}_i$) (pesos de frangos de corte)

Os resultados das comparações das diversas estruturas disponíveis na função `lme()` do R, para a matriz de covariâncias intra-indivíduos, são apresentados na Tabela 28. As comparações realizadas pelo teste da razão de verossimilhança e o critério AIC são favoráveis à escolha do modelo F6, que utiliza a matriz $\mathbf{R}_i = \text{UN}$ (não estruturada). O elevado número de parâmetros da estrutura UN (15 parâmetros) penalizou o valor do BIC, que foi favorável à escolha do modelo F7, que admite a estrutura ARH(1), tendo somente 6 parâmetros.

Tabela 28 - Estatísticas de ajuste dos modelos de F5 ao F9 (MV) (peso corporal de frangos)

Modelos	$\mathbf{R}_i$	gl	AIC	BIC	-2logVeross	Comparação	χ^2	p-valor
F5	VC	8	2418,790	2446,083	2402,790			
F6	UN	32	2295,926	2405,099	2231,926	F5×F6	170,86	<,0001
F7	ARH(1)	15	2310,253	2361,428	2280,253	F6×F7	48,32	0,0001
F8	AR(1)	9	2420,658	2451,363	2402,658	F6×F8	170,73	<,0001
F9	CS	9	2417,262	2447,967	2399,262	F6×F9	167,33	<,0001

O Quadro a seguir apresenta as estimativas dos parâmetros de efeito fixo do modelo F6.1, seus erros padrões, estatística t e p-valor, que indicam a importância de todos esses efeitos. Esse quadro foi obtido com o uso da função `summary(F6.1)`.

Coefficients:				
	Value	Std.Error	t-value	p-value
(Intercept)	-13.06203	3.214614	-4.063328	1e-04
tempo	137.32365	4.373606	31.398267	0e+00
I(tempo^2)	16.18157	0.908643	17.808490	0e+00
I(tempo^2):sexMacho	5.86523	0.853782	6.869700	0e+00

Com as estimativas dos parâmetros pode-se escrever as equações ajustadas para o peso corporal (em gramas) dos frangos de corte em função do tempo (em semanas), para cada um dos sexos:

$$y_{ij} = \begin{cases} -13,0620 + 137,3236t + 16,1816t_{ij}^2 & \text{se fêmea;} \\ -13,0620 + 137,3236t + 22,0468t_{ij}^2 & \text{se macho.} \end{cases} \quad (10)$$

O gráfico dessa equação é apresentados na Figura 18.

Figura 18 - Equação ajustada do modelo F6.1

Comparando a equação (10) com a obtida no estudo com enfoque multivariado de Lima (1988), verifica-se que as equações possuem a mesma estrutura de média, porém, por (10), somente o termo quadrático passa a ser diferente para os diferentes sexos.

Derivando-se a equação em relação a t_{ij} , obtém-se as equação que fornecem estimativas das taxas de crescimento de peso, quais sejam:

$$y'_{ij} = \begin{cases} 137,3236 + 32,3631t_{ij} & \text{se fêmea;} \\ 137,3236 + 44,0936t_{ij} & \text{se macho.} \end{cases}$$

4.3.2 Diagnóstico do modelo F6.1

Um diagnóstico informal do modelo F6.1 pode ser feito com base nos gráficos apresentados na Figura 19. Pode-se perceber no gráfico (a) uma distribuição aleatória dos resíduos ao redor do valor zero e no gráfico (b) uma confirmação da distribuição normal dos resíduos. Essas características indicam um bom ajuste do modelo F6.1.

Figura 19 - Diagnóstico do ajuste do modelo F6.1

4.4 Algumas comparações entre o R e o proc mixed do SAS

O `proc mixed` do SAS® e a função `lme()` do R têm uma sintaxe própria para especificar os modelos lineares mistos. Ambos os *softwares* requerem um arquivo de dados no formato univariado. As principais linhas de comando desses *softwares* que foram usadas no ajuste dos modelos estudados no presente trabalho podem ser consultadas nos anexos D ao I. Resumos dos resultados obtidos por esses *softwares* são apresentados nos anexos A, B e C.

A implementação do modelo A1 (α_{S1} -caseína) nos dois *softwares* fica:

R:

```
A1<-lme(alphas1CN~Dia+Grupo2.f+Dia:Grupo2.f,
        random=~Dia,method="REML", data=UATa2,
        control=lmeControl(opt="optim"))
```

SAS:

```
proc mixed data=UAT covtest;
  Title 'Modelo A1';
  class Rep Grupo;
  model alphas1CN=Dia Grupo Dia*Grupo / solution ddfm=sat;
  random int Dia / subject=Rep(Grupo) type=un g v solution;
Run;
```

A seguir serão feitos comentários sobre alguns aspectos importantes desses dois *softwares* utilizando, principalmente, o modelo A1.

Comparando os resultados obtidos, percebe-se que as estimativas dos parâmetros da parte fixa do modelo A1 (Tabela 29 do anexo A) são idênticas, embora haja discordância no número de graus de liberdade associado a algumas fontes de variação.

Os dois *softwares* utilizam métodos diferentes para calcular o número de graus de liberdade, mas o `proc mixed` permite uma maior flexibilidade na especificação do método para realizar este cálculo. Por exemplo: com a opção `ddfm=sat` obtém-se uma boa aproximação deste número pelo método de Satterthwaite.

Ao aplicar as funções `summary()` ou `anova()` no R para objetos ajustados pela função `lme()`, o número de graus de liberdade do denominador do teste-F (ou graus de liberdade do teste-t) associado aos efeitos fixos entre indivíduos, é calculado como:

$$n^{\circ}\text{indivíduos} - n^{\circ}\text{efeitos fixos entre indivíduos}$$

Para os efeitos fixos intra-indivíduos, o número de graus de liberdade é calculado como:

$$n^{\circ}\text{observações} - n^{\circ}\text{indivíduos} - n^{\circ}\text{efeitos fixos intra-indivíduos}$$

Por exemplo: no modelo A1 (Tabela 29 do anexo A), o número de graus de liberdade do teste-t para o fator entre indivíduos, Grupo, é igual a $15 - 3 = 12$. Já para o fator intra-indivíduo, Dia, tem-se $75 - 15 - 3 = 57$ graus de liberdade.

A execução de um teste da razão de verossimilhança comparando os modelos A1 e A4, por exemplo, é feita automaticamente no R com o comando:

```
>anova(A1,A4)
```

Resultando na seguinte mensagem de erro:

```
>anova(A1,A4)
Erro em anova.lme(A1, A4):
All fitted objects must have the same estimation method.
```

Esta mensagem de erro aparece porque os dois modelos foram ajustados utilizando dois métodos distintos - MV e MVR, respectivamente. A comparação entre os modelos A1 e A4.1, que foram ajustados pelo mesmo método (MVR), é feita pelo comando:

```
anova(A1,A4.1)
```

Resultando em:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
A1	1	10	296.2789	318.6200	-138.1395			
A4.1	2	6	276.8811	290.4572	-132.4406	1 vs 2	11.39778	0.0224
Warning message:								
In anova.lme(A1, A4.1) :								
Fitted objects with different fixed effects.								
REML comparisons are not meaningful.								

Observe que neste caso o teste é executado, mas é feita uma notificação de que os dois modelos comparados apresentam diferentes estruturas na parte fixa do modelo linear misto. As comparações entre modelos não são feitas de forma automática no `proc mixed` do SAS, devendo-se ficar atento a esses detalhes envolvidos na execução de um teste de razão de verossimilhança.

O número dos graus de liberdade associado aos modelos que estão sendo comparados é dado por:

$$gl = n^{\circ}\text{parâmetros de efeito fixo} \\ + n^{\circ}\text{número de parâmetros de covariâncias das matrizes } \mathbf{D} \text{ e } \mathbf{R}_i$$

Segundo West et al. (2007) quando é usada a função `anova()` no R, obtém-se um quadro de análise de variância com estatísticas de teste-F do tipo I, enquanto o `proc mixed` informa as estatísticas de teste-F do tipo III. Se for necessário, utiliza-se a opção `htype=1` para que se apresente o teste-F do tipo I.

As estimativas dos parâmetros de covariâncias diferem nos dois *softwares* e esta diferença pode ser devida a diferenças nos critérios de convergência e nos métodos de otimização.

Os valores de $-2\log Veross$ são diferentes nos dois *softwares* também devido ao uso de diferentes métodos de otimização. Os valores de AIC e BIC diferem nos dois *softwares* devido ao uso de diferentes fórmulas de cálculo. No `proc mixed` o AIC é calculado como:

$$-2\log Veross + 2(n^{\circ}\text{parâmetros de covariância})$$

Na função `lme()` do R o AIC é calculado como:

$$-2\log Veross + 2(n^{\circ}\text{efeitos fixos} + n^{\circ}\text{parâmetros de covariância})$$

De um modo geral, apesar dos valores de AIC e BIC não serem comparáveis entre os dois *softwares*, eles sempre conduzem à mesma conclusão, ou seja, sempre indicam o mesmo modelo como sendo o melhor. A opção `covtest` no `proc mixed` permite a execução do teste-z de

Wald para os parâmetros de covariância, baseado nas suas estimativas e nos erros padrões dessas estimativas. O uso deste teste é desaconselhável e deve ser substituído pelo teste da razão de verossimilhança, (WEST et al., 2007).

Durante o ajuste dos modelos A6 e B1, o `proc mixed` mostrou a seguinte nota:

Convergence criteria met but final hessian is not positive
definite.

Isto quer dizer que, embora o critério de convergência tenha sido cumprido, a matriz hessiana, que é utilizada para calcular os erros padrões das estimativas dos parâmetros de covariância, não é positiva definida. Segundo West et al. (2007), para evitar este problema recomenda-se o uso do algoritmo Newton-Raphson (N-R) ao invés do algoritmo *Fisher Scoring*.

Os algoritmos de otimização utilizados na estimação de um modelo linear misto, por MV ou MVR, devem assegurar que as estimativas das matrizes $\mathbf{D}$ e $\mathbf{R}_i$ sejam positivas definidas. No `proc mixed` se essa condição não for satisfeita na última iteração, é exibida a seguinte nota:

WARNING: Stopped because of infinite likelihood.

Neste caso é aconselhável escolher outro algoritmo de otimização ou mudar a estrutura de covariâncias assumida para a matriz $\mathbf{R}_i$ ou $\mathbf{D}$. No caso da função `lme()` do R a opção é o método de decomposição *log-Cholesky*, que proporciona uma grande simplificação nos problemas de otimização. Para maiores detalhes sobre o assunto, ver Pinheiro e Bates (1996).

5 CONCLUSÕES

Os resultados obtidos nessa pesquisa permitem extrair as seguintes conclusões:

- A abordagem proposta foi eficiente na indicação do modelo linear misto adequado a cada um dos conjuntos de dados. A qualidade e facilidade de obtenção dos gráficos do *software* R foram decisivas no estudo.
- Os resultados dos Testes da Razão da Verossimilhança e dos Critérios de Informação de Akaike - AIC e de Schwarz - BIC podem indicar escolhas de estruturas de covariâncias diferentes, havendo a necessidade de aprofundar os conhecimentos (via simulação, por exemplo) sobre o melhor critério de decisão.
- A função `lme()` do *software* R e o `proc mixed` do SAS são excelentes pacotes para análise de modelos lineares mistos, muito abrangentes, interativos e de fácil implementação. O *software* R tem o atrativo de realizar comparações entre modelos de forma automática.
- Desenvolvimento de novos métodos estatísticos de forma, muitas vezes, imediata a ser obtida pelos usuários do R, incentiva o desenvolvimento de pacotes práticos, porém a maior dificuldade encontrada é concentrar em um só pacote as técnicas de ajuste de modelos, o que permitiria ao usuário pleno controle do *software* sem se tornar um obstáculo durante a interação. Por exemplo, o pacote `lme()` é direcionado a modelos lineares mistos e também o pacote `lme4()` com a mesma abordagem e diferentes técnicas de especificações dos modelos.
- O SAS apresenta grande flexibilidade para o ajuste de modelos lineares mistos, com grande performance do `proc mixed`, permite assim uma especificação geral da matriz de covariâncias dos efeitos aleatórios **D** e do resíduo.
- Muitos usuários utilizam SAS em harmonia com o R afim de desencadear o poder combinado destes *softwares*. Relacionar uma maior quantidade de *softwares* para o estudo estatístico aponta alternativas melhores para soluções em diversos campos e estimula o desenvolvimento e aprimoramento da pesquisa.

REFERÊNCIAS

- BATES, D.; PINHEIRO, J.C. lmer for SAS PROC MIXED users. [s.l.: s.n.], 2000. Disponível em: <<http://biostat.hitchcock.org/FacultyandStaff/OnlineManuals/PDF%20Files/lmesas.pdf>>. Acesso em: 10 jan. 2008.
- BATES, D. Fitting linear mixed models in R. **The Newsletter of the R Project**, [s.l.], v. 5, p. 27-30, May 2005. Disponível em: <<http://cran.osmirror.nl/doc/Rnews/Rnews-2005-1.pdf#page=27>>. Acesso em: 15 mar. 2008.
- CAMARINHA FILHO, J.A. **Modelos lineares mistos: estruturas de matrizes de variâncias e co-variâncias e seleção de modelos**. 2002. 85 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.
- COSTA, S.C. **Modelos lineares generalizados mistos para dados longitudinais**. 2003. 110 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.
- CROWDER, M.J.; HAND, D.J. **Analysis of repeated measures**. London: Chapman & Hall, 1990. 257 p.
- DAVIDIAN, M.; GILTINAN, D.M. **Nonlinear models for repeated measurement data**. London: Chapman & Hall, 1995. 359 p.
- DIGGLE, P.J. An approach to the analysis of repeated measurements. **Biometrics**, Washington, v. 44, p. 959-971, 1988.
- DIGGLE, P.; HEAGERTY, P.; LIANG, K.Y; ZEGER, S. **Analysis of longitudinal data**. 2nd ed. Oxford: Oxford University Press, 2002. 379 p.
- EVERITT, B.S. **An R and S-Plus® companion to multivariate analysis**. London: Springer-Verlag, 2005. 221 p.
- FAUSTO, M.A.; CARNEIRO, M.; ANTUNES, C.M.F.; PINTO, J.A.; COLOSIMO, E.A. O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados. **Cadernos de Saúde Pública (FIOCRUZ)**, Rio de Janeiro, v. 24, n. 3, p. 513-524, mar. 2008.
- FERNANDES, A.M. **Efeitos dos níveis de células somáticas sobre a qualidade do leite integral obtido por processo UAT direto**. 2007. 109 p. Tese (Doutorado em Zootecnia) - Faculdade de Zootecnia e Engenharia de Alimentos, Pirassununga, 2007.
- FERNANDEZ, G. Model selection in PROC MIXED: a user-friendly SAS macro application. In: SAS GLOBAL FORUM, 2007. Orlando, 2007. p. 191. Disponível em: <<http://www2.sas.com/proceedings/forum2007/191-2007.pdf>>. Acesso em: 12 abr. 2009
- FLORIANO, E.P.; MÜLER, I.; FINGER, C.A.G.; SCHNEIDER, R. Ajuste e seleção de modelos tradicionais para série temporal de dados de altura de árvores. **Ciência Florestal**, Santa Maria, v. 16, n. 2, p. 177-199, 2006.
- FOX, J. “Linear mixed models” appendix to an R and S-Plus companion to applied regression. **Sage Publications**, May 2002. Disponível em: <<http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf>>. Acesso em: 7 set. 2007.

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

GURKA, M.J. Selecting the best linear mixed model under REML. **The American Statistician**, Washington, v. 60, n. 1, p. 20-26, Feb. 2006.

HELMS, R.W. Intentionally incomplete longitudinal designs: 1. Methodology and comparison of some full span designs. **Statistics in Medicine**, Chichester, v. 11, p. 1889-1913, 1992.

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

JENNRICH, R.I.; SCHLUCHTER, M.D. Unbalanced repeated measures models with structure covariance matrices. **Biometrics**, Washington, v. 42, p. 805-20, 1986.

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

LIMA, C.G. **Análise de curvas de crescimento de aves: um enfoque multivariado**. 1988. 69 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1988.

LIMA, C.G. **Análise de dados longitudinais provenientes de experimentos em blocos casualizados**. 1996. 119 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1996.

LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D. **SAS system for mixed models**. Cary: Statistical Analysis System Institute, 1996. 633 p.

LITTELL, R.C.; HENRY, P.R.; AMMERMAN, C.B. Statistical analysis of repeated measures data using SAS procedures. **Journal of Animal Science**, Albany, v. 76, n. 4, 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.

MOLENBERGHS G.; VERBEK, G. **Models for discrete longitudinal data**. New York: Springer-Verlang, 2005. 683 p.

NATIS, L. **Modelos lineares hierárquicos**. 2000. 87 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2000.

NGO, L.; BRAND, R. **Model selection in linear mixed effects models using SAS® PROC MIXED**. [s.l.: s.n.], 2002. Disponível em: <<http://www2.sas.com/proceedings/sugi22/STATS/PAPER284.PDF>>. Acesso em: 5 abr. 2007.

PERRI, S.H.V.; IEMMA, A.F. Procedure of software SAS® for the analysis of mixed models. **Scientia Agricola**, Piracicaba, v. 56, n. 4, 1999.

PINHEIRO, J.C. **Topics in mixed effects models**. 1994. 210 p. Thesis (PhD) - University of Wisconsin, Madison 1994.

PINHEIRO, J.C.; BATES, D.M. **Unconstrained parameterizations for variance-covariance matrices**. Wisconsin: University of Wisconsin, 1996. 6 p.

_____. **lme and nlme: mixed-effects methods and classes S and S-PLUS**. version 3.0. Wisconsin: Bell Labs, Lucent Technologies and University, June 1999. 32 p.

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

POTTHOFF, R.F.; ROY, S.N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrika**, Cambridge, v. 51, p. 313-26, 1964.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing** Viena: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 jan. 2009.

SANTORO, K.R.; BARBOSA, S.B.P.; BRASIL, L.H.A.; SANTOS, E.S. Descrição da variabilidade intra-específica em observações de peso-idade de bovinos por funções de covariância. **Archivos de Zootecnia**, Córdoba, v. 55, p. 251-262, 2006.

SAS INSTITUTE INC. **SAS OnlineDoc® 9.1.3**. Cary, 2004. Disponível em: <<http://support.sas.com/onlinedoc/913/docMainpage.jsp>>. Acesso em: 11 ago. 2008.

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

SCHEFFÉ, H. **The analysis of variance**. New York: John Wiley, 1959. 477 p.

THOMPSON, L.A. **R (and S-PLUS) manual to accompany agresti's categorical data analysis (2002)**. 2nd ed. 2008. 276 p. Disponível em: <<http://home.comcast.net/lthompson221/Splusdiscrete2.pdf>>. Acesso em: 20 jan. 2008.

UNIVERSITY OF CALIFORNIA. **Statistical computing**. Disponível em: <<http://www.ats.ucla.edu/stat/>>. Acesso em: 13 jan. 2008.

VENEZUELA, M.K. **Modelos lineares generalizados para análise de dados com medidas repetidas**. 2003. 95 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, São Paulo, 2003.

VERBEK, G.; MOLENBERGHS, G. **Linear mixed models in practice: a SAS-oriented approach**. Cary: SAS Institute, 1997. 306 p.

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

VONESH, E.F.; CHINCHILLI, V.M. **Linear an nonlinear models for the analisys of repeated measurements**. New York: Marcel Dekker, 1997. 560 p.

VIEIRA, F.T.P.A.; SILVA, J.A.A.; FERREIRA, R.L.C.; CRUZ, M.A.O.M.; FERRAZ, I. Uma abordagem multivariada em experimentos Silvipastoral com *Leucaena leucocephala* (Lam.) de Wit no Agreste de Pernambuco. **Ciência Florestal**, Santa Maria, v. 17, n. 4, p. 333-342, 2007.

WARE, J.H. Linear models for the analysis of longitudinal studies. **The American Statistician**, Washington, v. 39, p. 95-101, 1985.

WEST, B.T.; WELCH, K.B.; GALECKI, A.T. **Linear mixed models**: a practical guide using statistical software. New York: Chapman & Hall, 2007. 339 p.

WOLFINGER, R.D. Covariance structure selection in general mixed models. **Communications in Statistics**, New York, v. 22, p. 1079-1106, 1993.

XAVIER, L.H. **Modelos univariados e multivariados para análise de medidas repetidas e verificação da acúria do modelo univariado por meio de simulação**. 91 p. 2000. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2000.

ANEXOS

ANEXO A - Comparação dos resultados obtidos pelo proc mixed do SAS e função lme() do R (α_{S1} -caseína)

Tabela 29 - Resultados dos modelos A1 e A2 produzidos pelos proc mixed do SAS e função lme() do R (α_{S1} -caseína)

	proc mixed		função lme()	
Modelo A1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,4251	12	9,425058	57
β_1 Dia	-0,02249	12	-0,022491	57
β_2 Grupo2.f1	1,4959	12	1,495922	12
β_3 Grupo2.f2	1,2034	12	1,203449	12
β_4 Dia:Grupo2.f1	-0,01354	12	-0,013544	57
β_5 Dia:Grupo2.f2	-0,00878	12	-0,008783	57
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	0,7189		$9,51496 \times 10^{-01}$	
$\sigma_{\beta_0 t}$	0,006374		0,426 ^a	
σ_t^2	$5,544 \times 10^{-06}$		$4,990098 \times 10^{-05}$	
σ^2	1,2763		1,209738	
Critérios de Informação				
-2logVeross	275,8		276,27	
AIC	283,8		296,27	
BIC	286,7		318,6	
Modelo A2 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,4251	57	9,425058	57
β_1 Dia	-0,02249	24,1	-0,022491	57
β_2 Grupo2.f1	1,4959	57	1,495922	12
β_3 Grupo2.f2	1,2034	57	1,203449	12
β_4 Dia:Grupo2.f1	-0,01354	24,1	-0,013544	57
β_5 Dia:Grupo2.f2	-0,00878	24,1	-0,008783	57
Parâmetros de Covariância				
σ_t	0,006235		0,0002353654	
σ^2	1,5026		1,5025529211	
Critérios de Informação				
-2logVeross	285,3		285,263	
AIC	289,3		301,263	
BIC	290,7		319,1358	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

Tabela 30 - Resultados dos modelos A3, A3.1 e A4 produzidos pelos `proc mixed` do SAS e função `lme()` do R (α_{S1} -caseína)

	proc mixed		função lme()	
Modelo A3 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,4251	21	9,425058	57
β_1 Dia	-0,02249	57	-0,022491	57
β_2 Grupo2.f1	1,4959	21	1,495922	12
β_3 Grupo2.f2	1,2034	21	1,203449	12
β_4 Dia:Grupo2.f1	-0,01354	57	-0,013544	57
β_5 Dia:Grupo2.f2	-0,00878	57	-0,008783	57
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	1,5233		1,523285	
σ^2	1,2858		1,285771	
Critérios de Informação				
-2logVeross	277,9		277,857	
AIC	281,9		293,8571	
BIC	283,3		311,7299	
Modelo A3.1 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,4251	28.3	9,425058	57
β_1 Dia	-0,02249	60	-0,022491	57
β_2 Grupo2.f1	1,4959	28.3	1,495922	12
β_3 Grupo2.f2	1,2034	28.3	1,203449	12
β_4 Dia:Grupo2.f1	-0,01354	60	-0,013544	57
β_5 Dia:Grupo2.f2	-0,00878	60	-0,008783	57
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	1,1801		1,180055	
σ^2	1,2215		1,221482	
Critérios de Informação				
-2logVeross	254,3		254,292	
AIC	270,3		270,292	
BIC	276,0		288,8319	
Modelo A4 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,8835	19,4	9,883515	59
β_1 Dia	-0,02993	60	-0,029934	59
β_2 Grupo2.f1	0,6616	15	0,661600	12
β_3 Grupo2.f2	0,6624	15	0,662400	12
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	1,1673		1,167322	
σ^2	1,2851		1,285146	
Critérios de Informação				
-2logVeross	257,3		257,3404	
AIC	269,3		269,3404	
BIC	273,6		283,2454	

Tabela 31 - Resultados dos modelos A4.1, A5 e A6 produzidos pelos `proc mixed` do SAS e função `lme()` do R (α_{S1} -caseína)

	proc mixed		função lme()	
Modelo A4.1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,8835	14,9	9,883515	59
β_1 Dia	-0,02993	59	-0,029934	59
β_2 Grupo2.f1	0,6616	12	0,661600	12
β_3 Grupo2.f2	0,6624	12	0,662400	12
Parâmetros de Covariância				
$\sigma^2_{\beta_0}$	1,5191		1,519054	
σ^2	1,306		1,306928	
Critérios de Informação				
-2logVeross	264,9		264,8812	
AIC	268,9		276,8811	
BIC	270,3		290,4572	
Modelo A5 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,8459	13,1	9,845928	59
β_1 Dia	-0,02670	14	-0,026703	59
β_2 Grupo2.f1	1,3869	12	1,386824	12
β_3 Grupo2.f2	0,9421	12	0,942065	12
Parâmetros de Covariância				
σ^2	(-)		1,461227	
Critérios de Informação				
-2logVeross	244,5		244,521	
AIC	274,5		282,5215	
BIC	285,1		325,5124	
Modelo A6 (MVR)				
Mensagem de aviso	<i>Convergence criteria met but final hessian is not positive definite.^a</i>			
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,8835	15,69	9,883515	59
β_1 Dia	-0,02993	-9,12	-0,029934	59
β_2 Grupo2.f1	0,6616	0,78	0,661600	12
β_3 Grupo2.f2	0,6624	0,78	0,662400	12
Parâmetros de Covariância				
$\sigma^2_{\beta_0}$	1,5191		1,519054	
σ^2	1,3069		1,306928	
Critérios de Informação				
-2logVeross	244,5		264,8812	
AIC	274,5		278,8811	
BIC	285,1		294,7199	

Nota: Sinais convencionais utilizados:

(-) valor não calculado.

^a critério de convergência cumprido porém matriz hessiana não é positiva definida.

Tabela 32 - Resultados dos modelos A7 e A7.1 produzidos pelos `proc mixed` do SAS e função `lme()` do R (α_{S1} -caseína)

	proc mixed		função lme()	
Modelo A7 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	10,3248	28,1	10,324849	59
β_1 Dia	-0,02993	60	-0,029934	59
Parâmetros de Covariância				
$\sigma^2_{\beta_0}$	1,2647		1,264710	
σ^2	1,2851		1,285146	
Critérios de Informação				
-2logVeross	258,3		258,3324	
AIC	266,3		266,3325	
BIC	269,2		275,6024	
Modelo A7.1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	10,3248	25,6	10,324849	59
β_1 Dia	-0,02993	59	-0,029934	59
Parâmetros de Covariância				
$\sigma^2_{\beta_0}$	1,3690		1,369049	
σ^2	1,3069		1,306928	
Critérios de Informação				
-2logVeross	268,4		268,3582	
AIC	272,4		276,3582	
BIC	273,8		285,52	

ANEXO B - Comparação dos resultados obtidos pelo proc mixed do SAS e função lme() do R (β -caseína)

Tabela 33 - Resultados do modelo B1 produzidos pelos proc mixed do SAS e função lme() do R (β -caseína)

	proc mixed		função lme()	
Modelo B1 (MVR)				
Mensagem de aviso	<i>Convergence criteria met but final hessian is not positive definite.^a</i>			
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6092	12,5	9,609194	54
β_1 Dia	-0,02879	26	-0,028788	54
β_2 I(Dia)^2	0,000015	42	0,000015	54
β_3 Grupo2.f1	1,8712	12,5	1,871220	12
β_4 Grupo2.f2	1,9645	12,5	1,964542	12
β_5 Dia:Grupo2.f1	0,03248	26	0,032482	54
β_6 Dia:Grupo2.f2	-0,00336	26	-0,003359	54
β_7 I(Dia)^2:Grupo2.f1	-0,00032	42	-0,000323	54
β_8 I(Dia)^2:Grupo2.f2	-0,00001	42	-0,000013	54
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	0,4177		$4,909313 \times 10^{-1}$	
$\sigma_{\beta_0 t}$	-0,00211		-0,264 ^b	
$\sigma_{\beta_0 t^2}$	-0,00002		0,082 ^b	
σ_t^2	0,000521		$6,463625 \times 10^{-4}$	
σ_{tt^2}	$-1,31 \times 10^{-6}$		-0,866 ^b	
$\sigma_{t^2}^2$	$1,67 \times 10^{-6}$		$1,589845 \times 10^{-8}$	
σ^2	0,8792		$8,360083 \times 10^{-1}$	
Critérios de Informação				
-2logVeross	303,1		335,5160	
AIC	315,1		370,5504	
BIC	319,4		303,516	

Nota: Sinais convencionais utilizados:

^a critério de convergência cumprido porém matriz hessiana não é positiva definida.

^b valor de correlação.

Tabela 34 - Resultados dos modelos B2 e B3 produzidos pelos `proc mixed` do SAS e função `lme()` do R (β -caseína)

	proc mixed		função lme()	
Modelo B2 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6092	42	9,609194	54
β_1 Dia	-0,02879	39,1	-0,028788	54
β_2 I(Dia)^2	0,000015	33,2	0,000015	54
β_3 Grupo2.f1	1,8712	42	1,871220	12
β_4 Grupo2.f2	1,9645	42	1,964542	12
β_5 Dia:Grupo2.f1	0,03248	39,1	0,032482	54
β_6 Dia:Grupo2.f2	-0,00336	39,1	-0,003359	54
β_7 I(Dia)^2:Grupo2.f1	-0,00032	33,2	-0,000323	54
β_8 I(Dia)^2:Grupo2.f2	-0,00001	33,2	-0,000013	54
Parâmetros de Covariância				
σ_t^2	0,000732		$7,569532 \times 10^{-4}$	
σ_{tt^2}	$-3,53 \times 10^{-6}$		-0,924 ^a	
$\sigma_{t^2}^2$	$1,909 \times 10^{-8}$		$2,274313 \times 10^{-8}$	
σ^2	0,9754		$9,603675 \times 10^{-1}$	
Critérios de Informação				
-2logVeross	304,8		304,874	
AIC	312,8		330,874	
BIC	315,7		359,3395	
Modelo B3 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6092	54	9,609194	54
β_1 Dia	-0,02879	63.5	-0,028788	54
β_2 I(Dia)^2	0,000015	54	0,000015	54
β_3 Grupo2.f1	1,8712	54	1,871220	12
β_4 Grupo2.f2	1,9645	54	1,964542	12
β_5 Dia:Grupo2.f1	0,03248	63.5	0,032482	54
β_6 Dia:Grupo2.f2	-0,00336	63.5	-0,003359	54
β_7 I(Dia)^2:Grupo2.f1	-0,00032	54	-0,000323	54
β_8 I(Dia) ^ 2:Grupo2.f2	-0,00001	54	-0,000013	54
Parâmetros de Covariância				
σ_t^2	0,000214		0,0002143219	
σ^2	1,0487		1,0487411588	
Critérios de Informação				
-2logVeross	307,1		307,0544	
AIC	311,1		329,0544	
BIC	312,5		353,1406	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

Tabela 35 - Resultados dos modelos B3.1 e B4 produzidos pelos `proc mixed` do SAS e função `lme()` do R (β -caseína)

	proc mixed		função lme()	
Modelo B3.1 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6092	60	9,609194	54
β_1 Dia	-0,02879	70,1	-0,028788	54
β_2 I(Dia)^2	0,000015	60	0,000015	54
β_3 Grupo2.f1	1,8712	60	1,871220	12
β_4 Grupo2.f2	1,9645	60	1,964542	12
β_5 Dia:Grupo2.f1	0,03248	70,1	0,032482	54
β_6 Dia:Grupo2.f2	-0,00336	70,1	-0,003359	54
β_7 I(Dia)^2:Grupo2.f1	-0,00032	60	-0,000323	54
β_8 I(Dia)^2:Grupo2.f2	-0,00001	60	-0,000013	54
Parâmetros de Covariância				
σ_t^2	0,000168		0,0001675824	
σ^2	0,9439		0,9438670472	
Critérios de Informação				
-2logVeross	234,9		234,8892	
AIC	256,9		256,8893	
BIC	264,7		282,3817	
Modelo B4 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,3343	60	9,334318	56
β_1 Dia	-0,01449	74,9	-0,014486	56
β_2 I(Dia)^2	-0,00010	60	-0,000097	56
β_3 Grupo2.f1	2,6630	60	2,663005	12
β_4 Grupo2.f2	1,9974	60	1,997384	12
β_5 Dia:Grupo2.f1	-0,00872	29,1	-0,008718	56
β_6 Dia:Grupo2.f2	-0,00507	29,1	-0,005068	56
Parâmetros de Covariância				
σ_t^2	0,000166		0,0001658850	
σ^2	0,9898		0,9898059919	
Critérios de Informação				
-2logVeross	237,7		237,7408	
AIC	255,7		255,7407	
BIC	262,1		276,5981	

Tabela 36 - Resultados dos modelos B5, B6 e B6.1 produzidos pelos `proc mixed` do SAS e função `lme()` do R (β -caseína)

	proc mixed		função lme()	
Modelo B5 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,4521	71,9	9,452112	58
β_1 Dia	-0,01908	69,9	-0,019081	58
β_2 I(Dia)^2	-0,00010	59,7	-0,000097	58
β_3 Grupo2.f1	2,4395	71,2	2,439536	12
β_4 Grupo2.f2	1,8675	71,2	1,867472	12
Parâmetros de Covariância				
σ_t^2	0,000172		0,0001721791	
σ^2	0,9929		0,9929438347	
Critérios de Informação				
-2logVeross	238,4		238,3982	
AIC	252,4		252,3982	
BIC	257,4		268,6206	
Modelo B6 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6918	74,7	9,691750	59
β_1 Dia	-0,03146	28,6	-0,031464	59
β_3 Grupo2.f1	2,4364	70,8	2,436383	12
β_4 Grupo2.f2	1,8656	70,8	1,865639	12
Parâmetros de Covariância				
σ_t^2	0,000171		0,0001713829	
σ^2	1,0125		1,0124804187	
Critérios de Informação				
-2logVeross	239,6		239,5618	
AIC	251,6		251,5619	
BIC	255,8		265,4668	
Modelo B6.1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6859	70,8	9,685937	59
β_1 Dia	-0,03146	25,4	-0,031464	59
β_3 Grupo2.f1	2,4474	67,3	2,447410	12
β_4 Grupo2.f2	1,8720	67,3	1,872049	12
Parâmetros de Covariância				
σ_t^2	0,000194		0,0001943882	
σ^2	1,0550		1,0550274842	
Critérios de Informação				
-2logVeross	250,0		249,955	
AIC	254,0		261,9551	
BIC	255,4		275,5311	

Tabela 37 - Resultados dos modelos B7 e B8 produzidos pelos `proc mixed` do SAS e função `lme()` do R (β -caseína)

	proc mixed		função lme()	
Modelo B7 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	10,1767	12	10,176727	59
β_1 Dia	-0,03954	14	-0,039541	59
β_3 Grupo2.f1	2,1485	12	2,148538	12
β_4 Grupo2.f2	1,7136	12	1,713595	12
Parâmetros de Covariância				
σ^2	(-)		1,153887	
Critérios de Informação				
-2logVeross	220,6		220,5864	
AIC	250,6		258,5863	
BIC	261,2		301,5772	
Modelo B8 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	9,6941	13,7	9,694070	59
β_1 Dia	-0,03146	19	-0,031464	59
β_3 Grupo2.f1	2,4319	12	2,431981	12
β_4 Grupo2.f2	1,8630	12	1,863080	12
Parâmetros de Covariância				
σ_t^2	0,000149		0,0001491660	
σ^2	0,9113		1,3651597596	
Critérios de Informação				
-2logVeross	247,0		246,9538	
AIC	253,0		260,9537	
BIC	255,1		276,7925	

Nota: Sinais convencionais utilizados:

(-) valor não calculado.

ANEXO C - Comparação dos resultados obtidos pelo proc mixed do SAS e função lme() doR (peso corporal de frangos)

Tabela 38 - Resultados dos modelos F1 e F2 produzidos pelos proc mixed do SAS e função lme() do R (peso corporal de frangos)

	proc mixed		função lme()	
Modelo F1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-35,8791	158	-35,87912	188
β_1 tempo	149,205	62,9	149,20330	188
β_2 I(tempo)^2	15,9176	31,8	15,91758	188
β_3 sexMacho	-20,4968	158	-20,49682	30
β_4 tempo:sexMacho	6,3650	62,9	6,36500	188
β_5 I(tempo)^2:sexMacho	5,6620	31,8	5,66199	188
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	0		788,93156	
$\sigma_{\beta_0 t}$	-1143,62		-0,988 ^a	
$\sigma_{\beta_0 t^2}$	143,68		0,754 ^a	
σ_t^2	1702,53		1123,17500	
σ_{tt^2}	-210,76		-0,819 ^a	
$\sigma_{t^2}^2$	34,3708		25,39608	
σ^2	1451,77		1373,47762	
Critérios de Informação				
-2logVeross	2356,6		2368,098	
AIC	2368,6		2394,097	
BIC	2377,4		2438,096	
Modelo F2 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-35,8791	158	-35,87912	188
β_1 tempo	149,205	152	149,20330	188
β_2 I(tempo)^2	15,9176	72,4	15,91758	188
β_3 sexMacho	-20,4968	158	-20,49682	30
β_4 tempo:sexMacho	6,3650	152	6,36500	188
β_5 I(tempo)^2:sexMacho	5,6620	72,4	5,66199	188
Parâmetros de Covariância				
σ_t^2	336,76		336,75923	
σ_{tt^2}	-47,9624		-0,674 ^a	
$\sigma_{t^2}^2$	15,0228		15,02026	
σ^2	1451,65		1451,60722	
Critérios de Informação				
-2logVeross	2374,1		2374,092	
AIC	2382,1		2394,091	
BIC	2388,0		2427,936	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

Tabela 39 - Resultados dos modelos F2.1 e F3 produzidos pelos `proc mixed` do SAS e função `lme()` do R (peso corporal de frangos)

	proc mixed		função lme()	
Modelo F2.1 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-35,8791	160	-35,87912	188
β_1 tempo	149,205	162	149,20330	188
β_2 I(tempo)^2	15,9176	79,5	15,91758	188
β_3 sexMacho	-20,4968	160	-20,49682	30
β_4 tempo:sexMacho	6,3650	162	6,36500	188
β_5 I(tempo)^2:sexMacho	5,6620	79,5	5,66199	188
Parâmetros de Covariância				
σ_t^2	307,22		307,21971	
σ_{tt^2}	-43,5415		-0,668 ^a	
$\sigma_{t^2}^2$	13,8297		13,82745	
σ^2	1433,50		1433,46121	
Critérios de Informação				
-2logVeross	2401,5		2401,536	
AIC	2421,5		2421,535	
BIC	2436,2		2455,652	
Modelo F3 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-35,8791	120	-35,87912	188
β_1 tempo	149,205	168	149,20330	188
β_2 I(tempo)^2	15,9176	158	15,91758	188
β_3 sexMacho	-20,4968	120	-20,49682	30
β_4 tempo:sexMacho	6,3650	168	6,36500	188
β_5 I(tempo)^2:sexMacho	5,6620	158	5,66199	188
Parâmetros de Covariância				
$\sigma_{\beta_0}^2$	1137,66		1139,1931	
$\sigma_{\beta_0 t}$	-719,56		-0,944 ^a	
σ_t^2	509,93		509,7793	
σ^2	1882,14		1881,8449	
Critérios de Informação				
-2logVeross	2395,2		2395,168	
AIC	2403,2		2415,169	
BIC	2409,0		2449,014	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

Tabela 40 - Resultados dos modelos F4, F5 e F6 produzidos pelos `proc mixed` do SAS e função `lme()` do R (peso corporal de frangos)

	proc mixed		função lme()	
Modelo F4 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-40,6788	160	-40,67874	189
β_1 tempo	152,98	162	152,98251	189
β_2 I(tempo)^2	15,4484	79,5	15,44845	189
β_3 sexMacho	-12,4132	160	-12,41325	30
β_5 I(tempo)^2:sexMacho	6,4522	79,5	6,45211	189
Parâmetros de Covariância				
σ_t^2	308,11		308,11638	
σ_{tt^2}	-43,6738		-0,668 ^a	
$\sigma_{t^2}^2$	13,8490		13,84671	
σ^2	1434,80		1434,75886	
Critérios de Informação				
-2logVeross	2401,8		2401,75	
AIC	2419,8		2419,751	
BIC	2432,9		2450,456	
Modelo F5 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-48,0491	160	-48,04911	189
β_1 tempo	152,98	162	152,98251	189
β_2 I(tempo)^2	15,6036	84,1	15,60355	189
β_5 I(tempo) ^2:sexMacho	6,1908	32	6,19089	189
Parâmetros de Covariância				
σ_t^2	311,10		311,10068	
σ_{tt^2}	-44,1101		-0,67 ^a	
$\sigma_{t^2}^2$	13,9110		13,90885	
σ^2	1441,85		1441,80688	
Critérios de Informação				
-2logVeross	2402,8		2402,79	
AIC	2418,8		2418,790	
BIC	2430,5		2446,083	
Modelo F6 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-13,0719	32	-13,06212	189
β_1 tempo	137,33	32	137,32353	189
β_2 I(tempo)^2	16,1813	39	16,18149	189
β_5 I(tempo)^2:sexMacho	5,8659	32	5,86530	189
Parâmetros de Covariância				
σ^2	(-)		18,74201	
Critérios de Informação				
-2logVeross	2231,9		2231,926	
AIC	2295,9		2295,926	
BIC	2342,8		2405,099	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

(-) valor não calculados.

Tabela 41 - Resultados dos modelos F6.1 e F7 produzidos pelos `proc mixed` do SAS e função `lme()` do R (peso corporal de frangos)

	proc mixed		função lme()	
Modelo F6.1 (MVR)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-13,0719	31	-13,06212	189
β_1 tempo	137,33	31	137,32353	189
β_2 I(tempo)^2	16,1813	37,5	16,18149	189
β_5 I(tempo)^2:sexMacho	5,8659	30	5,86530	189
Parâmetros de Covariância				
σ^2	(-)		18,87966	
Critérios de Informação				
-2logVeross	2222,6		2222,618	
AIC	2278,6		2286,617	
BIC	2319,7		2395,213	
Modelo F7 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-7,5429	28,7	-11,34979	189
β_1 tempo	111,92	40,8	114,89775	189
β_2 I(tempo)^2	18,9821	63,6	18,27207	189
β_5 I(tempo)^2:sexMacho	-6,2142	30,7	5,71989	189
Parâmetros de Covariância				
σ_t^2	209,32		192,313789	
σ_{tt^2}	-17,1075		-0,742 ^a	
$\sigma_{t^2}^2$	6,5610		1,683245	
σ^2	(-)		110,891141	
Critérios de Informação				
-2logVeross	2278,3		2280,252	
AIC	2308,3		2310,253	
BIC	2330,3		2361,428	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

(-) valor não calculados.

Tabela 42 - Resultados dos modelos F8 e F9 produzidos pelos `proc mixed` do SAS e função `lme()` do R (peso corporal de frangos)

	proc mixed		função lme()	
Modelo F8 (MV)				
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	-47,6272	69,2	-47,63159	189
β_1 tempo	152,50	108	152,50115	189
β_2 I(tempo)^2	15,6819	80,3	15,68240	189
β_5 I(tempo)^2:sexMacho	-6,1941	31,9	6,19196	189
Parâmetros de Covariância				
σ_t^2	323,21		322,00629	
σ_{tt^2}	-46,2650		-0,68 ^a	
$\sigma_{t^2}^2$	14,3407		14,37851	
σ^2	1410,17		1411,06580	
Critérios de Informação				
-2logVeross	2402,7		2402,658	
AIC	2420,7		2420,658	
BIC	2433,8		2451,363	
Modelo F9 (MV)				
Mensagem de aviso	Stopped because of infinite likelihood. ^b			
Parâmetros Efeito Fixo	Valor	gl den	Valor	gl den
β_0 (intercepto)	(-)	(-)	-48,04911	189
β_1 tempo	(-)	(-)	152,98251	189
β_2 I(tempo)^2	(-)	(-)	15,62557	189
β_5 I(tempo)^2:sexMacho	(-)	(-)	6,15380	189
Parâmetros de Covariância				
σ_t^2	579,69 ^c		328,00303	
σ_{tt^2}	-73,4951 ^c		-0,664 ^a	
$\sigma_{t^2}^2$	17,0848 ^c		14,05871	
σ^2	1,5026 ^c		1244,05772	
Critérios de Informação				
-2logVeross	(-)		2399,262	
AIC	(-)		2417,262	
BIC	(-)		2447,967	

Nota: Sinais convencionais utilizados:

^a valor de correlação.

^b iteração: parou porque verossimilhança infinita.

^c valores calculados com base na última iteração.

(-) valor não calculados.

ANEXO D - Programa no *Software* R para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de leite UAT com relação a variável α_{S1} -caseína

```

require(nlme)
require(lattice)
UAT<-read.table("UHTV.txt",h=T)
Dia<-UAT$Dia
Grupo<-UAT$Grupo
alphas1CN<-UAT$alphas1CN
# novo código para o fator Grupo
Grupo2<-Grupo
Grupo2[Grupo==3]<-0
Grupo2[Grupo==2]<-2
Grupo2[Grupo==1]<-1
Grupo2.f<-factor(Grupo2)
lote<-c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5),
        rep(6,5),rep(7,5),rep(8,5),rep(9,5),rep(10,5),
        rep(11,5),rep(12,5),rep(13,5),rep(14,5),rep(15,5))
lote<-factor(lote)
dados<- subset(data.frame(UAT, Grupo2.f))
dados1<-subset(data.frame(UAT,Grupo2.f,lote))
#####
####          VARIÁVEL RESPOSTA alphas1-caseína          ####
#####
UATa<-groupedData(alphas1CN~Dia|Grupo/Rep, data =dados, order.groups=F)
UATa2<-groupedData(alphas1CN~Dia|lote, data =dados1, order.groups=F)
UAT.g1<-groupedData(alphas1CN~Dia|Grupo/Rep, outer=~Grupo, data=dados)
UAT.g1.1<-groupedData(alphas1CN~Dia|Grupo, order.groups=F, data=dados)
#-----
#ESTATÍSTICA DESCRITIVA
#-----
# Número de Observações para cada Dia
# Média de alphas1CN para cada Dia por Grupo
tapply(UAT$alphas1CN, list(Grupo,factor(UAT$Dia)), mean, na.rm=TRUE)
# desvio padrão de alphas1CN para cada dia por Grupo
tapply(UAT$alphas1CN, list(Grupo,factor(UAT$Dia)), sd, na.rm=TRUE)
# Valores mínimos de alphas1CN para cada Dia
tapply(UAT$alphas1CN, factor(UAT$Dia), min, na.rm=TRUE)
# Valores máximos de alphas1CN para cada Dia
tapply(UAT$alphas1CN, factor(UAT$Dia), max, na.rm=TRUE)
#####
####          MODELOS LINEARES MISTOS AJUSTADOS          ####
#####
#MODELO A1 (MVR)
#-----
A1<-lme(alphas1CN~Dia+Grupo2.f+Dia:Grupo2.f,
        random=~Dia,method="REML", data=UATa2,
        control=lmeControl(opt="optim"))

```

```

summary(A1)
#MODELO A2 (MVR)
#-----
A2<-lme(alphas1CN~Dia+Grupo2.f+Dia:Grupo2.f,
        random=~Dia-1, method="REML", data=UATa2)
summary(A2)
#MODELO A3 (MVR)
#-----
A3<-update(A1,random=~1)
summary(A3)
#ANOVA=====
anova(A1,A2,A3)
#MODELO A3.1 (MV)
#-----
A3.1<-update(A3, method="ML")
#MODELO A4 (MV)
#-----
A4<-update(A3.1,fixed=~Dia +Grupo2.f)
summary(A4)
#ANOVA=====
anova(A3.1,A4)
#MODELO A4.1 (MVR)
#-----
A4.1<-update(A4,method="REML")
#ANOVA=====
anova(A4.1)
#####
###          ESTRUTURAS DE COVARIÂNCIA - MATRIZ Ri          ####
#####
# FUNÇÃO gls() - Mínimos quadrados generalizados
#MODELO A5-Não Estruturada
#-----
A5<-gls(alphas1CN~Dia+Grupo2.f,data=UATa2,
        corr=corSymm(form=~1|lote),
        weights=varIdent(form=~1|Dia),
        control=list( maxlter=100))
summary(A5)
#MODELO A6-Simetria Composta (CS)
#-----
A6<-update(A4.1, corr=corCompSymm(,form=~1|lote),
          control=lmeControl(opt="optim"))
summary(A6)
#ANOVA=====
anova(A4.1,A5,A6)
# Modelo A7 (MV)
#-----
A7<-update(A4.1, fixed=~Dia, method="ML")
# Modelo A7.1 (MVR)
#-----

```

```

A7.1<-update(A7, method="REML")
summary(A7.1)
#####
####                                GRÁFICOS                                ####
#####
# dados individuais por grupo=====
plot(UAT.g1, display=1,outer=T, key=FALSE, xlab = "Tempo(Dia)",
ylab="alphas1CN (mg/mL)",main="Dados individuais por Grupo")
# perfil médio por grupo=====
plot(UAT.g1.1, display = "CCS", aspect = 2, key = F, xlab = "Tempo(Dia)",
ylab="alphas1CN (mg/mL)",main = "Perfil Médio por Grupo")
#boxplot do coeficiente de regressão=====
g1.list <-lmList(alphas1CN ~ Dia|lote,
subset=Grupo=="1",data=UATa2)
g2.list <- lmList(alphas1CN ~Dia|lote,
subset=Grupo=="2", data=UATa2)
g3.list <- lmList(alphas1CN ~Dia|lote,
subset = Grupo=="3", data=UATa2)
g1.coef <- coef(g1.list)
g2.coef <- coef(g2.list)
g3.coef<- coef(g3.list)
X11()
old <- par(mfrow=c(1,2))
boxplot(g1.coef[,1], g2.coef[,1],g3.coef[,1], main="Interceptos",
names=c("Grupo 1", "Grupo 2","Grupo 3"))
boxplot(g1.coef[,2], g2.coef[,2],g3.coef[,2], main="Inclinação",
names=c("Grupo 1", "Grupo 2","Grupo 3"))
par(old)
# valores observados e curva ajustada pelo modelo=====
plot(augPred(A1),layout=c(5,3),between=list(y=c(1,1)))
# valores ajustado x residuos padronizados=====
plot(A7.1,id=0.05,adj=-0.03, abline=0)
# Q-Q plot dos resíduos=====
qqnorm(A7.1,id=0.05,adj=-0.04)
# valores ajustados x alphas1CN=====
plot(A7.1, alphas1CN~fitted(.),id=0.05, adj=-0.3)
#Q-Q plot dos efeitos aleatórios=====
qqnorm (A7.1, ~ ranef(.), id = 0.10)
# gráfico dos parametros fixos do modelo=====
#g1: (B0+B3) + B1t + B2T^2
curve(10.324849 -0.029934*x ,0, 120, xlab="Dia",
      ylab="yij marginal predito", lty=3, ylim=c(5,15),lwd=2)
      legend(locator(1),c(expression(10.324849-0.029934*t)),
      lty=c(3,2,1), lwd=c(2,2,2))
#gráfico de envelope modelo A7.1=====
require(car)
qq.plot(resid(A7.1),pch=19,col="black")

```

ANEXO E - Programa no *Software* SAS para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de leite UAT com relação a variável α_{S1} -caseína

```

options nodate nocenter ps=10000 ls=100;
infile "c:\dados\UAT.prn";
input Dia CCS alphas1CN alphas2CN betaCN kappaCN;
if CCS le 316 then Grupo=1; else if CCS > 316 and
CCS le 560 then Grupo=2; else Grupo=3;
proc print data=UAT;
Run;

proc mixed data=UAT covtest;
  Title 'Modelo A1';
  class Rep Grupo;
  model alphas1CN = Dia Grupo Dia*Grupo / solution ddfm=sat;
  random int Dia / subject=Rep(Grupo) type=un g v solution;
Run;

proc mixed data=UAT covtest;
  Title 'Modelo A2';
  class Grupo Rep;
  model alphas1CN = Dia Grupo Dia*Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) g solution;
Run;

proc mixed data=UAT covtest method=REML;
  Title 'Modelo A3';
  class Rep Grupo;
  model alphas1CN = Dia Grupo Dia*Grupo / solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
Run;

proc mixed data=UAT covtest method=ML;
  Title 'Modelo A3.1';
  class Rep Grupo;
  model alphas1CN = Dia Grupo Dia*Grupo / solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
Run;

proc mixed data=UAT covtest method=ML;
  Title 'Modelo A4';
  class Rep Grupo;
  model alphas1CN = Dia Grupo / solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
Run;

proc mixed data=UAT covtest method=REML;
  Title 'Modelo A4.1';

```

```

class Rep Grupo;
model alphas1CN = Dia Grupo / solution ddfm=sat;
random int / subject=Rep(Grupo) g v solution;
Run;

```

```

proc mixed data=UAT covtest method=REML;
  Title 'Modelo A5';
  class Rep Grupo;
  model alphas1CN = Dia Grupo / solution ddfm=sat;
  repeated / subject = Rep(Grupo) type = un r rcorr;
Run;

```

```

proc mixed data=UAT covtest method=REML;
  Title 'Modelo A6';
  class Rep Grupo;
  model alphas1CN = Dia Grupo / noint solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
  repeated / subject = Rep(Grupo) type=cs r rcorr;
Run;

```

```

proc mixed data=UAT covtest method=ML;
  Title 'Modelo A7';
  class Rep Grupo;
  model alphas1CN = Dia / solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
Run;

```

```

proc mixed data=UAT covtest method=REML;
  Title 'Modelo A7.1';
  class Rep Grupo;
  model alphas1CN = Dia / solution ddfm=sat;
  random int / subject=Rep(Grupo) g v solution;
Run;

```

ANEXO F - Programa no *Software* R para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de leite UAT com relação a variável β -caseína

```
rm(list=ls(all=TRUE))
require(nlme)
require(lattice)
UAT<-read.table("UHTV.txt",h=T)
Dia<-UAT$Dia
Grupo<-UAT$Grupo
betaCN<-UAT$betaCN
# novo código para o fator Grupo
Grupo2<-Grupo
Grupo2[Grupo==3]<-0
Grupo2[Grupo==2]<-2
Grupo2[Grupo==1]<-1
Grupo2.f<-factor(Grupo2)
lote<-c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5),
        rep(6,5),rep(7,5),rep(8,5),rep(9,5),rep(10,5),
        rep(11,5),rep(12,5),rep(13,5),rep(14,5),rep(15,5))
lote<-factor(lote)
dados<-subset(data.frame(UAT, Grupo2.f))
dados1<-subset(data.frame(UAT,Grupo2.f,lote) )
#-----
#ESTATÍSTICA DESCRITIVA
#-----
# Média de betaCN para cada Dia por Grupo
tapply(UAT$betaCN, list(Grupo,factor(UAT$Dia)), mean, na.rm=TRUE)
# desvio padrão de betaCN para cada dia por Grupo
tapply(UAT$betaCN, list(Grupo,factor(UAT$Dia)), sd, na.rm=TRUE)
#####
####          VARIÁVEL RESPOSTA beta-caseína          ##
#####
UATa<-groupedData(betaCN~Dia|Grupo/Rep, data =dados, order.groups=F)
UATa2<-groupedData(betaCN~Dia|lote, data =dados1)
UAT.g1<-groupedData(betaCN~Dia|Grupo/Rep, outer=~Grupo, data=dados)
UAT.g1.1<-groupedData(betaCN~Dia|Grupo, order.groups=F, data=dados)
#####
###          MODELOS AJUSTADOS lme()          ####
#####
#MODELO B1 (MVR)
#-----
B1<-lme(betaCN~Dia+I(Dia^2)+Grupo2.f+Dia:Grupo2.f+I(Dia^2):Grupo2.f,
        random=~Dia+I(Dia^2),method="REML",
        data=UATa2)
summary(B1)
#MODELO B2 (MVR)
#-----
B2<-lme(betaCN~Dia+I(Dia^2)+Grupo2.f+Dia:Grupo2.f+I(Dia^2):Grupo2.f,
```

```

        random=~Dia+I(Dia^2)-1, method="REML", data=UATa2)
#MODELO B3 (MVR)
#-----
B3<-update(B2,random=~Dia-1)
summary(B3)
#ANOVA=====
anova(B1,B2,B3)
# MODELO B3.1 (MV)
#-----
B3.1<-update(B3,method="ML")
summary(B3.1)
anova(B3.1)
#MODELO B4 (MV)
#-----
B4<-update(B3.1, fixed=~Dia+I(Dia^2)+Grupo2.f+Dia:Grupo2.f)
summary(B4)
#MODELO B5 (MV)
#-----
B5<-update(B4, fixed=~Dia+I(Dia^2)+Grupo2.f)
summary(B5)
#MODELO B6 (MV)
#-----
B6<-update(B5, fixed=~Dia+Grupo2.f)
summary(B6)
#ANOVA=====
anova(B3.1,B4,B5,B6)
#MODELO B6.1(MVR)
#-----
B6.1<-update(B6, method="REML")
summary(B6.1)
#####
####          ESTRUTURAS DE COVARIÂNCIA - MATRIZ Ri          ####
#####
# FUNÇÃO GLS - Mínimos quadrados generalizados
#MODELO B7-Não Estruturada
#-----
B7<-glS(betaCN~Dia+Grupo2.f,data=UATa2,
        corr=corSymm(form=~1|lote),
        weights=varIdent(form=~1|Dia),
        control=list( maxIter=100))
summary(B7)
#MODELO B8-Simetria Composta (CS)
#-----
B8<-update(B6.1, corr=corCompSymm(,form=~1|lote),
          control=lmeControl(opt="optim"))
summary(B8)
#-----
B9<-glS(betaCN~Dia+Grupo1,data=UATa2,
        corr=corSymm(form=~1|lote),

```

```

weights=varIdent(form=~1|Dia),
control=list( maxlter=100), method="ML")
#-----
B9.1<-gls(betaCN~Dia,data=UATa2,
          corr=corSymm(form=~1|lote),
          weights=varIdent(form=~1|Dia),
          control=list( maxlter=100), method="REML")
#####
###                                GRÁFICOS                                ####
#####
# Dados individuais por grupo=====
plot(UAT.g1, display=1,outer=T, key=FALSE, xlab = "Tempo(Dia)",
ylab="betaCN (mg/mL)",main="Dados individuais por Grupo")
# Perfil médio por Grupo=====
interaction.plot(Dia, Grupo, UAT$betaCN, trace.label="Grupo",
xlab="Tempo (Dias)", ylab = "betaCN (mg/mL)")
# gráficos diagnóstico de regressão linear simples=====
par(mfrow=c(2,2))
plot(B)
# boxplot dos resíduos=====
bwplot(getGroups(UATa2,level=2)~resid(B))
#intervalos de confiança efeitos aleatórios=====
B.lis<-lmList(betaCN~Dia,data=UATa2)
plot(intervals(B.lis))
# gráfico de comparação - curva de dois modelos=====
plot(comparePred(B1,B6.1), lty=c(2,1),layout=c(5,3),
      between=list(y=c(0,0.5,0)))
#Equação para parâmetros de efeito fixo modelo B7=====
curve((10.176727 +2.148538 )-0.039541 *x, 8, 120, xlab = "Dia",
      ylab=expression(beta-caseina~mg/mL ), lty=3, ylim=c(5,15), lwd=2)
curve((10.176727 +1.713595 )-0.039541*x, 8, 120, add=T, lty = 2, lwd=2)
#g3: B0 + B1t
curve(10.176727 -0.039541*x, 8, 120, add=T, lty=1, lwd=2)
legend(locator(1),c("Grupo 1", "Grupo 2",
"Grupo 3"), lty=c(3,2,1), lwd=c(2,2,2))
#gráfico modelo B9.1=====
qq.plot(resid(B9.1),pch=19,col="black")

```


ANEXO G - Programa no *Software* SAS para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de leite UAT com relação a variável β -caseína

```

options nodate nocenter ps=10000 ls=100;
infile "c:\dados\UAT.prn";
input Dia CCS alphas1CN alphas2CN betaCN kappaCN;
if CCS le 316 then Grupo=1; else if CCS > 316 and
CCS le 560 then Grupo=2; else Grupo=3;
proc print data=UAT;
Run;
proc mixed data=UAT covtest;
  Title 'Modelo B1';
  class Rep Grupo;
  model betaCN = Dia Dia*Dia Grupo Dia*Grupo Dia*Dia*Grupo / solution ddfm=sat;
  random int Dia Dia*Dia / subject=Rep(Grupo) type=un g v solution;
Run;

proc mixed data=UAT covtest;
  Title 'Modelo B2';
  class Grupo Rep;
  model betaCN = Dia Dia*Dia Grupo Dia*Grupo Dia*Dia*Grupo / solution ddfm=sat;
  random Dia Dia*Dia / subject=Rep(Grupo) type=un g v solution;
Run;

proc mixed data=UAT covtest;
  Title 'Modelo B3';
  class Rep Grupo;
  model betaCN = Dia Dia*Dia Grupo Dia*Grupo Dia*Dia*Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) type=un g v solution;
Run;

proc mixed data=UAT covtest method=ML;
  Title 'Modelo B3.1';
  class Rep Grupo;
  model betaCN = Dia Dia*Dia Grupo Dia*Grupo Dia*Dia*Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) g v solution;
Run;

proc mixed data=UAT covtest method=ML;
  Title 'Modelo B4';
  class Rep Grupo;
  model betaCN = Dia Dia*Dia Grupo Dia*Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) g v solution;
Run;

proc mixed data=UAT covtest method=ML;
  Title 'Modelo B5';
  class Rep Grupo;

```

```

model betaCN = Dia Dia*Dia Grupo / solution ddfm=sat;
random Dia / subject=Rep(Grupo) g v solution;
Run;

```

```

proc mixed data=UAT covtest method=ML;
  Title 'Modelo B6';
  class Rep Grupo;
  model betaCN = Dia Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) g v solution;
Run;

```

```

proc mixed data=UAT covtest method=REML;
  Title 'Modelo B6.1';
  class Rep Grupo;
  model betaCN = Dia Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) g v solution;
Run;

```

```

proc mixed data=UAT covtest;
  Title 'Modelo B7';
  class Rep Grupo;
  model betaCN = Dia Grupo / solution ddfm=sat;
  repeated / subject = Rep(Grupo) type = un r rcorr;
Run;

```

```

proc mixed data=UAT covtest;
  Title 'Modelo B8';
  class Rep Grupo;
  model betaCN = Dia Grupo / solution ddfm=sat;
  random Dia / subject=Rep(Grupo) type=un g v solution;
  repeated / subject = Rep(Grupo) type =cs r rcorr;
Run;

```

ANEXO H - Programa no *Software* R para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de frango com relação a variável peso corporal em gramas

```

library(nlme)
frango<-read.table("fran.txt",head=T)
names(frango)
sex<-frango$sex
sex<-factor(c(rep(0,91),rep(1,133)))
frango<-subset(data.frame(frango, sex))
tempo=frango$tempo
peso=frango$peso
indiv=factor(frango$indiv)
frango<-groupedData(peso~tempo|indiv, data=frango, outer=~sex,
                    labels=list(x="tempo (semana)", y="Peso corporal (g)"))
#####
####                                MODELOS AJUSTADOS lme()                                ####
#####
#MODELO F1 (MVR)
#-----
F1<-lme(fixed=peso~tempo+I(tempo^2)+sex+tempo:sex+I(tempo^2):sex,
        random=~tempo+I(tempo^2),data=frango,control=lmeControl(opt="optim"))
summary(F1)
#MODELO F2 (MVR)
#-----
F2<-update(F1, random=~tempo+I(tempo^2)-1,data=frango)
summary(F2)
#MODELO F2.1 (MV)
#-----
F2.1<-update(F2, method="ML")
anova(F2, test="Wald")
summary(F2)
anova(F1,F2)
#MODELO F3 (MVR)
#-----
F3<-update(F1,random=~tempo)
summary(F3)
anova(F1,F2,F3)
#MODELO F4 (MV)
#-----
F4<-update(F2.1, fixed=peso~tempo+I(tempo^2)+sex+I(tempo^2):sex)
summary(F4)
#MODELO F5 (MV)
#-----
F5<-update(F4,fixed=peso~tempo+I(tempo^2)+I(tempo^2):sex)
summary(F5)
#####
###                                ESTRUTURAS DE COVARIÂNCIA - MATRIZ Ri                                ####
#####

```

```

# FUNÇÃO GLS - Mínimos quadrados generalizados
#MODELO F6-Não Estruturada
#-----
F6<-gls(peso~tempo+I(tempo^2)+I(tempo^2):sex,data=frango,
        corr=corSymm(form=~1|indiv),
        weights=varIdent(form=~1|tempo),
        control=list( maxlter=100),method="ML")

summary(F6)
#MODELO F6.1 (MVR)-Não Estruturada
#-----
F6.1<-gls(peso~tempo+I(tempo^2)+I(tempo^2):sex,data=frango,
        corr=corSymm(form=~1|indiv),
        weights=varIdent(form=~1|tempo),
        control=list( maxlter=100),method="REML")

#MODELO F7-Autoregressiva com heterogeneidade variancias ARH(1)
#-----
F7<-update(F5, corr=corAR1(form=~1|indiv),
        weight=varIdent(form=~1|tempo),
        control=lmeControl( opt="optim"))

summary(F7)
#MODELO F8-Autoregressiva AR(1)
#-----
F8<-update(F5, corr=corAR1(form=~tempo|indiv),
        control=lmeControl(opt="optim"))

summary(F8)

#MODELO F9-Simetria Composta (CS)
#-----
F9<-update(F5, corr=corCompSymm(,form=~1|indiv),
        control=lmeControl(opt="optim"))

summary(F9)
anova(F5,F6,F7,F8,F9)
#####
###                                GRÁFICOS                                ###
#####
#gráfico peso em gramas para cada indivíduo=====
plot(frango, layout=c(8,4), between=list(y=c(0,0.5,0)))
#perfil médio de resposta (fêmea e macho)=====
interaction.plot(frango$tempo,sex,frango$peso, ylim=c(90,2000), lty=c(1,10),
        ylab="Peso (g)", xlab="Tempo (semana)" )
#boxplot dos dados originais=====
boxplot(frango$peso~frango$tempo:frango$sex, las=3)
#diagnósticos gráficos=====
par(mfrow=c(3,2))
plot(tempo,resid(F))
plot(F) #which=4
# distribuição dos resíduos modelo F1=====
plot(F1,indiv~resid(.),abline=0)
#grafico equacoes modelo F6.1 =====

```

```

curve(-13.06203+137.32365 *x+16.18157 *x^2, 1, 7, xlab="Tempo (Semanas)",
      ylab= "peso (gramas)", lty = 3, ylim=c(50,1500), lwd = 2)
curve(-13.06203+137.32365*x+(16.18157+5.86523)*x^2,1,7,add=T, lty=2, lwd=2)
legend(locator(1),c("Femea", "Macho"), lty=c(3,2,1), lwd=c(2,2,2))
#grafico equacoes modelo F7 =====
curve(-11.34979+114.89775*x+18.27207*x^2, 1, 7, xlab="Tempo (Semanas)",
      ylab= "peso (gramas)", lty = 3, ylim=c(50,1500), lwd = 2)
curve(-11.34979+114.89775*x+(18.27207+5.71989)*x^2,1,7,add=T, lty=2, lwd=2)
legend(locator(1),c("Femea", "Macho"), lty=c(3,2,1), lwd=c(2,2,2))
#resíduos padronizados x valores ajustados modelo F6.1=====
plot(F6.1, resid(.,type="p")~fitted(.)|sex,
     layout=c(2,1), aspect=2, id=0.05, adj=-0.03, abline=0)
#resíduos padronizados x valores ajustados modelo F7=====
plot(F7, resid(.,type="p")~fitted(.)|sex,
     layout=c(2,1), aspect=2, id=0.05, adj=-0.03, abline=0)
#normalização dos resíduos modelo F6.1=====
qqnorm(F6.1,~resid(.)|sex)
#normalização dos resíduos modelo F7=====
qqnorm(F7,~resid(.)|sex)
#ajuste do modelo F6.1=====
require(car)
qq.plot(resid(F6.1),pch=19,col="black")

```

ANEXO I - Programa no *Software* SAS para verificar o desempenho do modelo linear misto por meio dos critérios de AIC, BIC para os dados de frango com relação a variável peso corporal em gramas

```

options nodate nocenter ps=10000 ls=100;
infile "c:\dados\frango.prn";
data frangoU (Keep=indiv sex tempo peso)
input indiv sex tempo peso;
proc print data=frangoU;
Run;

proc mixed data= frangoU info;
Title 'Modelo F1';
class sex;
model peso= tempo tempo*tempo sex tempo*sex tempo*tempo*sex /
      solution ddfm=sat;
random int tempo tempo*tempo/ subject=indiv type=un g v solution;
Run;

proc mixed data= frangoU info method=REML;
Title 'Modelo F2';
class sex;
model peso= tempo tempo*tempo sex tempo*sex tempo*tempo*sex /
      solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
Run;

proc mixed data= frangoU info method=ML;
Title 'Modelo F2.1';
class sex;
model peso= tempo tempo*tempo sex tempo*sex tempo*tempo*sex /
      solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
Run;

proc mixed data= frangoU info method=REML ;
Title 'Modelo F3';
class sex;
model peso= tempo tempo*tempo sex tempo*sex tempo*tempo*sex /
      solution ddfm=sat;
random int tempo / subject=indiv type=un g v solution;
Run;

proc mixed data= frangoU info method=ML;
Title 'Modelo F4';
class sex;
model peso= tempo tempo*tempo sex tempo*tempo*sex / solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
Run;

```

```
proc mixed data= frangoU info method=ML;
Title 'Modelo F5';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
Run;
```

```
proc mixed data= frangoU info method=ML ;
Title 'Modelo F6';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
repeated / subject = indiv type = un r rcorr;
Run;
```

```
proc mixed data= frangoU info method=REML ;
Title 'Modelo F6.1';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
*random int tempo/ type=un subject=indiv g v solution;
repeated / subject = indiv type = un r rcorr;
Run;
```

```
proc mixed data= frangoU info method=ML;
Title 'Modelo F7';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
repeated / subject = indiv type= ARH(1) r rcorr;
Run;
```

```
proc mixed data= frangoU info method=ML;
Title 'Modelo F8';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
repeated / subject = indiv type = AR(1) r rcorr;
Run;
```

```
proc mixed data= frangoU info method=ML ;
Title 'Modelo F9';
class sex;
model peso= tempo tempo*tempo tempo*tempo*sex / solution ddfm=sat;
random tempo tempo*tempo/ subject=indiv type=un g v solution;
repeated / subject = indiv type = CS r rcorr;
Run;
```