

**USO DE MODELO MISTO PARA A ANÁLISE
DE DADOS LONGITUDINAIS DE UM
EXPERIMENTO COM BOVINOS EM
LACTAÇÃO**

STEPHANIA MARIA TEIXEIRA FURTADO

2009

STEPHANIA MARIA TEIXEIRA FURTADO

**USO DE MODELO MISTO PARA A ANÁLISE DE DADOS
LONGITUDINAIS DE UM EXPERIMENTO COM BOVINOS EM
LACTAÇÃO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Paulo César Lima

LAVRAS
MINAS GERAIS – BRASIL
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Furtado, Stephania Maria Teixeira.

Uso de modelo misto para a análise de dados longitudinais de um
experimento com bovinos em lactação / Stephania Maria Teixeira
Furtado. – Lavras : UFLA, 2009.

46 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2009.

Orientador: Paulo César Lima.

Bibliografia.

1. Dados longitudinais. 2. Estrutura de covariância. 3. Modelo
misto. 4. Curva de lactação. I. Universidade Federal de Lavras. II.
Título.

CDD – 519.537
519.538

STEPHANIA MARIA TEIXEIRA FURTADO

**USO DE MODELO MISTO PARA A ANÁLISE DE DADOS
LONGITUDINAIS DE UM EXPERIMENTO COM BOVINOS EM
LACTAÇÃO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 31 de julho de 2009

Prof. Dr. Paulo César Lima	UFLA
Prof. Dr. Renato Ribeiro de Lima	UFLA
Prof. Dr. Agostinho Roberto de Abreu	UFLA
Prof. Dr. Rilke Tadeu Fonseca de Freitas	UFLA
Pesquisador Dr. Ary Ferreira de Freitas	EMBRAPA

Prof. Dr. Paulo César Lima
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

À memória de:

meu avô, José Furtado Ferreira.

OFEREÇO.

À minha mãe, Maria da Graça Furtado de Andrade, pelo amor, carinho, confiança e dedicação sem medida, que permitiram a realização deste trabalho.

Ao meu pai, Antônio Dante Teixeira de Andrade, por ser meu espelho.

Aos amigos, presença constante em minha vida.

DEDICO.

“Não acredite no que teus olhos te dizem, porque tudo o que eles vêem é limitação. Então, olha com entendimento, descobre o que já sabes, e verás como voar.”

Richard Bach

AGRADECIMENTOS

A Deus, fonte inesgotável de amor e paz.

Ao professor Paulo César Lima pelos ensinamentos, orientação, paciência, apoio e amizade desde a época da graduação.

Ao professor Renato Ribeiro de Lima, pela dedicação, apoio e estímulos fundamentais na conclusão deste trabalho.

À amiga Taciana Villela Savian pela paciência, apoio e presença sempre que precisei.

À FAPEMIG pela concessão da bolsa, fundamental na realização deste trabalho.

À Universidade Federal de Lavras e ao Programa de Pós-graduação em Estatística e Experimentação Agropecuária, bem como a todos os professores do Departamento de Exatas.

A todos os colegas da Pós-graduação do DEX, em especial à Patrícia Neves, pela grande amizade.

Às amigas Maria, Selminha, Edila, Josi (P) e Josi (G), Joyce e Vânia, pelo carinho, dedicação, apoio e paciência durante todos esses anos.

A todos os meus amigos, especialmente Cleber, Bella, Fogo, Luan, Tatá, Taty, Amanda, Kézia e Denise que estiveram comigo durante esse percurso.

A todos que contribuíram para a realização deste trabalho.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	iii
RESUMO	iv
ABSTRACT	v
1 INTRODUÇÃO	1
2 REVISÃO BIBLIOGRÁFICA	2
2.1 Metodologias usuais para análise de dados longitudinais	2
2.2 Estruturas de covariâncias	7
2.3 Experimentos de lactação	11
2.3.1 Modelo não-linear de Wood	12
2.3.2 Identidade de modelos	13
3 MATERIAL E MÉTODOS	15
3.1 Modelos estatísticos utilizados	15
4 RESULTADOS E DISCUSSÃO	19
4.1 Modelo misto	20
4.2 Modelo fixo	26
4.2.1 Efeito do tempo na produção de cada tratamento.....	29
5 CONCLUSÕES	34
REFERÊNCIAS BIBLIOGRÁFICAS	35
ANEXOS	40

LISTA DE TABELAS

TABELA	Páginas
TABELA 1	Número de observações/vaca para cada dieta..... 20
TABELA 2	Número de observações (n), médias diárias, desvios-padrão (DP) e valores mínimos e máximos para a produção de leite, em kg, de cada dieta..... 20
TABELA 3	Número de observações (n), médias diárias, desvios-padrão (DP) e valores mínimos e máximos para a produção de leite, em kg, de cada dieta para o conjunto de dados com um menor grau de desbalanceamento..... 21
TABELA 4	Número de observações (n), médias diárias, desvios-padrão (DP) e valores mínimos e máximos para a produção de leite, em kg, de cada dieta para o conjunto de dados balanceados 21
TABELA 5	Crítérios de informação de Akaike (AIC) e Schwarz (BIC) para os conjuntos de dados com um menor grau de desbalanceamento e balanceados 23
TABELA 6	Análise de variância para os efeitos fixos do modelo com estrutura da matriz Σ do tipo AR (1) para o conjunto de dados com um menor grau de desbalanceamento 24
TABELA 7	Análise de variância para os efeitos fixos do modelo com a estrutura de matriz Σ do tipo AR (1) para o conjunto de dados balanceados 24
TABELA 8	Equações ajustadas e coeficientes de determinação (R^2) dos modelos..... 26

TABELA 9	Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para o conjunto de dados com todas as lactações.....	26
TABELA 10	Critérios de informação para o ajuste do modelo fixo.....	27
TABELA 11	Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para o conjunto de dados com um menor grau de desbalanceamento.....	28
TABELA 12	Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para os dados balanceados.....	29
TABELA 13	Equações ajustadas e coeficientes de determinação (R^2) do modelo completo.....	29
TABELA 14	Equações estimadas para o conjunto de dados completo.....	31
TABELA 15	Hipóteses avaliadas e valores da estatística do teste F para o conjunto de dados com todas as lactações.....	32

LISTA DE FIGURAS

FIGURA	Páginas
FIGURA 1 Gráfico de caixas da produção de leite (kg) para cada dieta.....	19
FIGURA 2 Gráfico de caixas da produção de leite (kg) para cada dieta para o conjunto de dados com um menor grau de desbalanceamento.....	22
FIGURA 3 Gráficos de caixas para produção de leite (kg) de cada dieta quando o conjunto de dados é balanceado.....	22
FIGURA 4 Curvas de lactação ajustadas para os conjuntos de dados completo, com um menor grau de desbalanceamento e balanceado.....	25
FIGURA 5 Curvas de lactação das duas dietas para o conjunto de dados com todas as lactações	30

RESUMO

FURTADO, Stephania Maria Teixeira. **Uso de modelo misto para a análise de dados longitudinais de um experimento com bovinos em lactação**. 2009. 46p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.¹

Dados longitudinais podem fornecer informações sobre variações globais e individuais ao longo do tempo. A característica distinta de estudos longitudinais é a dimensão ordenada com que os dados são coletados e o fato de que as observações repetidas para um indivíduo tendem a ser correlacionadas. O objetivo desse trabalho foi o de analisar um conjunto de dados longitudinais obtidos em um experimento de lactação com bovinos. O experimento foi conduzido com o objetivo de se comparar dois tratamentos que consistiram de suplementação alimentar diferenciada (concentrado até 60 dias pós-parto e concentrado até 120 dias após o parto). Cada tratamento foi aplicado a 8 vacas, que foram agrupadas segundo produção no parto anterior, composição genética, peso, idade e ordem de parto. As análises foram realizadas utilizando-se o software SAS, através do ajuste de dois modelos estatísticos, um modelo considerando aleatório o efeito de blocos (modelo misto) e outro com componentes de efeitos fixos (parcelas subdivididas no tempo). Além disso, foi ajustado o modelo não-linear de Wood e utilizado o teste de identidade de modelos para a comparação dos modelos ajustados. Verificou-se que houve diferença entre os resultados para as duas metodologias, sendo que o ajuste do modelo misto foi adequado para esse conjunto de dados de lactação.

¹ **Comitê orientador:** Paulo César Lima – UFLA (Orientador), Renato Ribeiro de Lima – UFLA.

ABSTRACT

FURTADO, Stephania Maria Teixeira. **Mixed model to analyse experimental longitudinal data with dairy cattle**. 2009. 46p. Dissertation (Master in Statistics and Agricultural Experimentation) – Federal University of Lavras, Lavras, MG.²

Longitudinal data offer information about global and individual variation over time. The size of ordered collection and the relationship between the observations is a distinctive feature of longitudinal data. This work aimed to study longitudinal data from dairy cattle experiment, conducted to compare two treatments consisted of supplementation differentiated (ration until 60 and ration until 120 days postpartum). Each treatment was applied to eight cows, which were grouped according with previous production, genetic composition, weight, age and order of birth. The tests were performed using the software SAS by considering adjustment of two models, one by considering the block effects as random (mixed model) and another with fixed components (split-plot in time). Also, it was adjusted Wood nonlinear model, which were compared by models identity test. It was found that there was difference between results for both methodologies and the mixed model appropriate to analyse this lactation data.

² **Guidance committee:** Paulo César Lima – UFLA (Adviser), Renato Ribeiro de Lima –UFLA.

1 INTRODUÇÃO

A expressão “medidas repetidas” é usada para designar medidas feitas no mesmo indivíduo ou na mesma unidade experimental em mais de uma ocasião. Quando as observações estão dispostas em uma escala ordenada, ou seja, são obtidas sequencialmente, os dados são ditos longitudinais.

Dados longitudinais podem fornecer informações sobre variações globais e individuais ao longo do tempo. A razão para a tomada de dados longitudinais em experimentos é a suspeita de que os efeitos dos tratamentos em uma sequência de tempo incorporem informações relevantes da variação individual.

A estrutura da matriz de variâncias e covariâncias dos erros para dados longitudinais é muito variável em função do tipo de relação entre as observações. Assim, as análises estatísticas para esses dados dependem de condições especiais, devendo considerar as diferentes formas de correlação temporal entre os dados.

Para a análise estatística de dados longitudinais, são sugeridas diferentes metodologias, tais como: análise univariada (modelo fixo e misto), análise multivariada, regressão aleatória, entre outras.

Este trabalho teve por objetivo analisar um conjunto de dados longitudinais obtidos em um experimento de lactação com bovinos através do ajuste de um modelo misto, considerando aleatório o efeito de blocos. Objetivou-se também, a título de comparação, o ajuste de um modelo com componentes de efeitos fixos (parcelas subdivididas no tempo).

2 REVISÃO BIBLIOGRÁFICA

Existem basicamente duas formas de obtenção de medidas repetidas: uma correspondente à atribuição aleatória da ordem com que as mensurações são realizadas e outra, onde as observações são obtidas sequencialmente. No último caso, os dados são ditos longitudinais (Rocha, 2004).

Estudos com medidas repetidas têm a vantagem de requerer um número menor de unidades amostrais quando comparados com estudos do tipo transversal, que consistem na observação da variável resposta uma única vez em cada unidade amostral (Venezuela, 2003).

Segundo Rocha (2004), são duas as principais desvantagens de estudos com medidas repetidas: a primeira é que a análise é mais complexa, por requerer que as correlações das medidas realizadas no mesmo indivíduo sejam modeladas. A segunda está relacionada com a presença de dados incompletos ou omissos, o que, na prática, ocorre com frequência.

A característica distinta de estudos longitudinais é a dimensão ordenada com que os dados são coletados e o fato de que as observações repetidas para um indivíduo tendem a ser correlacionadas. Nesses casos, as respostas de tempos mais próximos são, em geral, mais fortemente correlacionadas que as de tempos mais distantes (Littell et al., 1996; Xavier, 2000).

2.1 Metodologias usuais para análise de dados longitudinais

Os Modelos de Regressão Aleatória (MRA) foram propostos como alternativa na avaliação genética de dados longitudinais (Henderson Júnior, 1982), como a produção de leite, curvas de crescimento e, atualmente, têm sido muito empregados na análise de dados de lactação. El Faro & Albuquerque (2003) e Araújo et al. (2006) apresentaram trabalhos descrevendo a variação genética da produção de leite utilizando modelos de regressão aleatória.

Dorneles (2006) também utilizou MRA para o estudo da curva e da persistência da lactação de vacas.

Liang & Zeger (1986) propõem analisar dados com medidas repetidas utilizando modelos lineares generalizados (MLGs), que foram propostos por Nelder & Weddeburn (1972). Costa (2003) cita que uma das formas de se levar em conta a correlação dos dados é modelar explicitamente a estrutura de correlação, utilizando-se a abordagem de equações de estimação generalizada (EEG), dadas por Liang & Zeger (1986). Esse método permite modelar a variabilidade entre as observações incluindo na análise a matriz de covariâncias dos erros (Σ). Os modelos lineares generalizados têm somente um componente aleatório, mas podem ser estendidos para ter efeitos aleatórios no preditor linear. Essa extensão é conhecida como modelos lineares generalizados mistos.

Costa et al. (2004) modela dados de proporções observados ao longo do tempo utilizando quatro modelos: modelo em parcelas subdivididas no tempo; modelo logístico, incorporando um fator de dispersão; modelo logístico, incorporando a matriz Σ e modelo considerando um fator de dispersão e um efeito aleatório. Eles concluíram que a inclusão do efeito aleatório e do parâmetro de dispersão constante, conjuntamente, explica melhor o comportamento dos dados.

Araújo (2007) propõe a análise de dados longitudinais de simulação expressos em forma de contagem utilizando a distribuição Poisson multivariada para modelar as dependências entre as diferentes medidas longitudinais em uma mesma unidade experimental. Ela concluiu que, na maioria dos casos, as simulações indicaram eficiência do método, embora a metodologia proposta necessite que as covariâncias sejam iguais, o que mostra a necessidade de estender os estudos de modo a permitir a utilização de estruturas de covariâncias menos restritivas.

Outra opção de análise para dados longitudinais é a análise multivariada de perfis (Morrison, 1990). Lima (1996) salienta que o enfoque multivariado permite a adoção de um modelo bastante geral para representar a matriz Σ , admitindo que as variâncias das respostas em cada tempo e as covariâncias das respostas entre tempos distintos sejam diferentes, desde que a variável resposta tenha distribuição normal.

Experimentos com medidas repetidas são análogos a experimentos em esquema de parcelas subdivididas em muitos aspectos. Uma vez que as medidas repetidas são tomadas na mesma unidade experimental durante um determinado período de tempo, a análise dos dados provenientes desses experimentos é frequentemente conduzida como um experimento em parcelas subdivididas no tempo.

De acordo com Xavier (2000), pode-se realizar a análise de dados longitudinais sob o enfoque de parcelas subdivididas no tempo de forma usual, sendo feitas pressuposições de que, tanto o erro da parcela, que engloba o fator de tratamentos ou grupos, como o erro da subparcela, onde são alocados os tempos e a interação entre tempos e tratamentos, tenham distribuição normal, sejam independentes e identicamente distribuídos, com variâncias constantes.

Steel & Torrie (1980) sugerem que a análise de dados longitudinais seja feita seguindo um modelo em parcelas subdivididas, tendo o efeito de tempo como subparcela. Entretanto, Huynh & Feldt (1970) mostraram que, em um delineamento de parcelas subdivididas com medidas repetidas no tempo, o teste F tem distribuição F exata com relação à parcela, mas com relação à subparcela, só terá distribuição F exata se a matriz Σ satisfizer a condição de esfericidade, isto é, que não haja heterogeneidade de variâncias entre diferentes observações da subparcela e que o comportamento da covariância entre elas seja constante.

Para verificar a condição de esfericidade da matriz Σ , Mauchly (1940), citado pelos autores Morrison (1990) e Xavier (2000), desenvolveu um teste

que verifica se uma população normal multivariada apresenta variâncias iguais e correlações nulas.

Esse teste utiliza a condição de esfericidade da matriz Σ nos indivíduos requeridos nos (t-1) contrastes ortogonais normalizados, em que t é o número de tempos avaliados. Existe mais de um conjunto com (t-1) contrastes ortogonais, sendo que um contraste ortogonal será normalizado quando for dividido pela sua norma Euclidiana.

A condição de esfericidade para as covariâncias dos contrastes é:

$$C_{(t-1) \times t} \Sigma_{(t \times t)} C'_{t \times (t-1)} = \lambda I_{(t-1) \times (t-1)},$$

em que:

- C é a matriz de coeficientes dos contrastes ortogonais normalizados que apresenta o total de hipóteses nulas;
- Σ é a matriz de covariâncias dos erros;
- λ é um escalar maior que zero;
- I é uma matriz identidade.

Satisfeita essa condição, a matriz Σ é dita esférica. A estatística de teste formulada por Mauchly (1940) é:

$$W = \frac{(t-1)^{(t-1)} |CSC'|}{(\text{tr}(CSC'))^{(t-1)}},$$

em que:

- S é a matriz de covariâncias amostrais para o erro intra-indivíduos, com v graus de liberdade;
- C é a matriz de coeficientes dos contrastes ortogonais normalizados;

- t é o número de medidas repetidas.

Para melhor acurácia dessa aproximação pela distribuição Qui-quadrado, foi definido o seguinte fator escalar:

$$\gamma = v - \frac{2t^2 - 3t + 3}{6(t-1)}.$$

Então a hipótese de que a matriz Σ satisfaz a condição de esfericidade, ou seja, hipótese $H_0: \mathbf{C}\Sigma\mathbf{C}' = \lambda\mathbf{I}$, pode ser verificada da seguinte forma:

$$\chi^2 = -\gamma \ln(W),$$

que tem distribuição χ^2 , com $f = \frac{1}{2}t(t-1)$ graus de liberdade. Quando $-\gamma \ln(W) > \chi_{\alpha, f}^2$, rejeita-se a hipótese de nulidade ao nível α de confiança.

Caso a condição de esfericidade não seja satisfeita, alguns autores, como Box (1954), Geisser & Greenhouse (1958) e Huynh & Feldt (1976), propuseram um fator (ε) para a correção dos graus de liberdade referentes aos efeitos de tempo e da interação entre tratamentos e tempo, possibilitando que esses possam ser usados de forma aproximada. Geisser & Greenhouse (1959) apresentaram um estimador para (ε), dado por:

$$\varepsilon = \frac{[\text{tr}(\mathbf{CSC}')]^2}{(t-1) \text{tr}(\mathbf{CSC}')^2}.$$

Box (1954) mostrou que $(t-1)^{-1} \leq \varepsilon \leq 1$, isto é, o valor máximo de 1 para ε , corresponde à condição de esfericidade. À medida que a matriz Σ se

afasta desse padrão, o valor de ε decresce até atingir $(t-1)^{-1}$, seu limite inferior, provocando, assim, a redução máxima no número de graus de liberdade.

Outra opção de análise para dados longitudinais é o ajuste de modelos mistos que contemplam efeitos fixos e aleatórios (Freitas, 2007). Segundo Riboldi (2007), a maior flexibilidade dos modelos mistos possibilita que seja contemplada a estrutura de interdependência das observações e permite acomodar a estrutura de correlação presente entre as medidas repetidas tomadas na mesma unidade experimental.

Em um experimento com cana-de-açúcar, Freitas (2007) analisa dados de medidas repetidas utilizando um modelo misto e modelando a matriz Σ . Rocha (2004) propõe estratégias para seleção das estruturas da matriz Σ em estudos com medidas repetidas, utilizando modelos mistos.

2.2 Estruturas de covariâncias

Quando a mesma unidade experimental é observada ao longo do tempo, espera-se que haja uma correlação entre essas unidades (Costa, 2003). Assim, em dados longitudinais, a matriz Σ não apresenta a estrutura pressuposta na análise usual de modelos de delineamentos de experimentos ($I\sigma^2$), existindo uma estrutura diferente para essa matriz.

Segundo Malheiros (1999), em geral, as análises estatísticas de dados longitudinais dependem de condições especiais para a matriz Σ , que, muitas vezes, não são consideradas. A estrutura de análise utilizada por muitos pesquisadores é grandemente simplificada em dados dessa natureza. As simplificações são: análise considerando um modelo fatorial, como se os dados fossem completamente aleatorizados e análise desconsiderando a correlação de erros induzida pelas medidas repetidas (Gill, 1986).

Uma consequência imediata de se ignorar a correlação existente entre as observações é que a significância aparente da diferença entre médias de

tratamentos é grosseiramente exagerada e a sensibilidade dos testes para interação é seriamente reduzida. Quando essa correlação é ignorada, as inferências podem ou não ser distorcidas, dependendo do grau de homogeneidade das variâncias e covariâncias dos dados nos diferentes tempos (Gill, 1986). Segundo Costa (2003), é essencial modelar de forma adequada a matriz Σ para que as inferências sobre as médias sejam válidas.

A correlação entre as mensurações no indivíduo ao longo do tempo pode ser modelada por meio de uma estrutura de covariâncias de erros (Rosário et al., 2005). De acordo com Rocha (2004), o modelo para essa matriz depende da maneira pela qual as observações foram obtidas e do conhecimento sobre o mecanismo gerador das observações.

Muitas são as estruturas de covariância que podem ser identificadas nos conjuntos de dados longitudinais e Xavier (2000) apresenta alguns exemplos dessas estruturas:

- i) Componentes de Variância: caracterizada por variâncias iguais e correlações nulas.

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

- ii) Simetria Composta: caracterizada por variâncias homogêneas e covariâncias constantes entre quaisquer observações.

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

- iii) Auto-regressiva de 1ª ordem - AR(1): apresenta variâncias homogêneas e correlações que diminuem exponencialmente à medida em que aumenta o intervalo de tempo entre as medidas repetidas. Denota-se por ρ o parâmetro auto-regressivo, de forma que, para um processo estacionário, assume-se que $|\rho| < 1$.

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

- iv) Toeplitz: similar a estrutura AR(1), mas com correlações variáveis à medida em que as distâncias entre tempos crescem.

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma^2 & \sigma_{12} & \sigma_{13} \\ & & \sigma^2 & \sigma_{12} \\ & & & \sigma^2 \end{bmatrix};$$

- v) Auto-regressiva heterogênea: caracterizada pela desigualdade de variâncias e covariâncias e pela maior correlação entre avaliações adjacentes.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho \\ & & & \sigma_4^2 \end{bmatrix},$$

- vi) Não-estruturada: todas as variâncias e covariâncias podem ser desiguais. Especifica uma matriz completamente geral, parametrizada em termos de variâncias e covariâncias. As variâncias são restritas a valores não negativos e as covariâncias não têm restrições.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix}.$$

Para a escolha adequada da melhor estrutura, é necessário utilizar algum critério de seleção. Dentre esses critérios, destacam-se o Critério de informação de Akaike (AIC) (Akaike, 1973) e o Critério Bayesiano de Schwarz (BIC) (Akaike, 1977) que são, na verdade, valores para os logaritmos das funções de verossimilhanças penalizadas pelo número de parâmetros estimados. Os dois critérios podem ser expressos por:

$$AIC = -2l + 2p$$
$$BIC = -2l + p \log n$$

em que l é o máximo do logaritmo da função de verossimilhança, p é o número de parâmetros do modelo e n é o número de observações.

A estrutura de covariâncias com valores do critério mais próximos de zero é considerada mais adequada aos dados.

2.3 Experimentos de lactação

O Brasil possui um grande rebanho bovino leiteiro, principalmente no Estado de Minas Gerais. Entretanto, os níveis de produtividade estão bastante aquém do desejado. Deficiências de manejo associadas à baixa qualidade genética do rebanho contribuem diretamente para esse resultado (Gonçalves et al., 2002).

Experimentos de lactação são importantes por possibilitarem uma previsão da produção total de leite. Estimar a produção total de leite de um animal é fundamental na tomada de decisões sobre manejo alimentar, reprodutivo, descarte de animais ou, até mesmo, em programas de seleção. Considerando-se essa estimativa, é possível a seleção de animais com base em sua lactação ou na de seus descendentes (Ferreira, 2003).

A representação gráfica da produção de leite em função do tempo durante uma lactação completa é denominada curva de lactação (Cobuci et al., 2000). Uma curva de lactação típica é composta de três fases principais: a produção inicial, o rápido crescimento da produção nos 30 a 90 dias pós-parto culminando no chamado pico de lactação e a fase de declínio, cuja taxa de redução denomina-se persistência na lactação (Ferreira, 2003).

Existem vários trabalhos que procuram identificar uma função matemática que melhor se ajusta à produção de leite de bovinos (Sherchand et

al., 1995). Um dos primeiros a propor um modelo para descrever a curva de lactação foi Brody et al. (1923). Novos modelos foram propostos por outros autores, dentre eles Wood (1967), que propôs a função gamma incompleta. Ali & Schaeffer (1987) estudaram um modelo de regressão múltipla. Mais recentemente métodos Bayesianos foram utilizados por Groenewald & Viljoen (2003). Ferreira (2003) utilizou modelos de regressão aleatória para ajustar curvas de lactação.

2.3.1 Modelo não-linear de Wood

Dentre os muitos modelos propostos, o modelo desenvolvido por Wood (1967) -- baseado na função gama incompleta -- tem sido preferido no ajuste de curvas de lactação, por apresentar poucos parâmetros e estes apresentarem interpretação biológica (Silva, 2003).

O modelo não-linear de Wood (1967) é dado por:

$$y_t = at^b \exp(-ct) + \varepsilon_t,$$

em que:

- y_t representa a produção de leite do animal no tempo de lactação t ;
- a é a produção inicial de leite;
- b é a taxa crescente de produção até o pico;
- c é a taxa de declínio da produção após o pico;
- ε_t é o erro aleatório considerado aditivo para estes dados, com $\varepsilon \sim NID(0, \sigma^2)$.

2.3.2 Identidade de modelos

O ajuste de equações para explicar a produção de leite é uma ferramenta essencial em estudos de lactação. Em casos experimentais, curvas de lactação podem ser ajustadas para cada tratamento e, através de testes estatísticos, pode-se verificar se os modelos ajustados são idênticos. Denomina-se identidade de modelos um conjunto de equações que pode ser representado por uma equação comum (Regazzi & Silva, 2004).

O teste para identidade de modelos de regressão e igualdade de parâmetros verifica a possibilidade de se realizar um único ajuste do modelo selecionado para um conjunto de observações e se existe semelhança entre os parâmetros isoladamente, ou entre diferentes combinações dos mesmos.

Neter & Wasserman (1990) testaram se duas equações de regressão linear simples eram idênticas, utilizando o teste F. Graybill (1976) apresentou um método geral para testar a hipótese de igualdade de um conjunto de modelos lineares empregando o teste F. Bates & Watts (1988) apresentaram um teste assintótico, baseado na razão de verossimilhança, para comparar modelos de regressão não linear, com aproximação dada pela estatística F.

Regazzi (1993) considerou o ajustamento de H equações de regressão polinomial de grau k , mediante o emprego da técnica de polinômios ortogonais e apresentou um método para testar as seguintes hipóteses: (a) H_0 : as H equações são idênticas; (b) H_0 : as H equações têm uma constante de regressão comum; (c) H_0 : As H equações têm um ou mais coeficientes de regressão iguais. Ele concluiu que o método apresentado é geral e pode ser usado em modelos polinomiais de qualquer grau, ortogonais ou não, e também em modelos de regressão múltipla.

Regazzi (1999) apresentou um método para testar as mesmas hipóteses citadas anteriormente, considerando dados provenientes de delineamentos experimentais. Considerando o caso de dados com repetições, Regazzi & Silva

(2004) apresentaram o teste da razão de verossimilhança para testar identidade de modelos de regressão não linear com aproximação dada pelas estatísticas χ^2 e F.

3 MATERIAL E MÉTODOS

Os dados utilizados neste trabalho são provenientes de um experimento conduzido na EMBRAPA Gado de Leite, localizada em Coronel Pacheco, Minas Gerais. O arquivo original continha 4350 registros de produção de leite, coletados diariamente entre os meses de outubro de 2000 a agosto de 2001, em um rebanho de 16 vacas. O número de observações variou com tempo de lactação de cada vaca, caracterizando um número variado de medidas repetidas no tempo por animal, sendo que o maior número de medidas repetidas foi de 310 e o menor 206.

Esse experimento foi conduzido com o objetivo de comparar duas dietas que consistiram de suplementação alimentar diferenciada: T_1 = concentrado até 60 dias pós-parto e T_2 = concentrado até 120 dias após o parto, quando a produção foi superior a 10Kg de leite/dia, fornecido na base de 1kg de concentrado (18% PB e 70% NDT) para 2kg de leite. O delineamento experimental foi o de blocos casualizados, sendo que cada dieta foi aplicada a 8 vacas, agrupadas segundo produção no parto anterior, composição genética, peso, idade e ordem de parto. As vacas foram mantidas em piquetes de capim-elefante, utilizando 30 dias de descanso e 3 dias de ocupação, sendo ordenhadas mecanicamente sem bezerro.

3.1 Modelos estatísticos utilizados

As análises foram realizadas utilizando-se o software estatístico SAS (SAS, 2003), sob as seguintes modelagens: análise univariada através do ajuste de um modelo misto e de um modelo fixo (esquema de parcelas subdivididas no tempo). Além disso, foi ajustado o modelo não-linear de Wood (1967) e utilizado o teste de identidade de modelos para a comparação dos modelos ajustados.

O ajuste do modelo misto foi feito através do PROC MIXED. O modelo ajustado foi:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + e_{ijk},$$

com $i = 1, \dots, I$, $j = 1, \dots, J$ e $k = 1, \dots, K$, em que:

- y_{ijk} é a produção de leite;
- μ é a constante inerente a todas observações;
- α_i é o efeito fixo de dieta;
- β_j é o efeito aleatório de bloco;
- γ_k é o efeito fixo de tempo;
- $(\alpha\gamma)_{ik}$ é o efeito fixo da interação entre dieta e tempo;
- e_{ijk} é o erro aleatório, em que $e \sim N(0, \Sigma)$.

As estruturas de covariância denominadas componentes de variância, simetria composta, não-estruturada, auto-regressiva de 1ª ordem, toeplitz e auto-regressiva heterogênea foram consideradas na análise utilizando o PROC MIXED. A matriz de covariâncias dos erros mais adequada foi selecionada pelo critério de informação de Akaike (AIC) (Akaike, 1973) e o critério Bayesiano de Schwarz (BIC) (Akaike, 1977).

A análise no esquema de parcelas subdivididas no tempo foi realizada através do procedimento PROC GLM, tendo o efeito de tempo na subparcela. O modelo estatístico utilizado foi:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + e_{ijk},$$

em que:

- y_{ijk} é o valor observado para a produção de leite no k -ésimo tempo para a i -ésima dieta no j -ésimo bloco;
- μ é a constante inerente a todas as observações;
- α_i é o efeito da i -ésima dieta;
- β_j é o efeito do j -ésimo bloco;
- $(\alpha\beta)_{ij}$ é o efeito devido a interação da i -ésima dieta com o j -ésimo bloco;
- γ_k é o efeito do k -ésimo tempo observado;
- $(\alpha\gamma)_{ik}$ é o efeito da interação entre a i -ésima dieta com o k -ésimo tempo;
- $(\beta\gamma)_{jk}$ é o efeito da interação entre o j -ésimo bloco com o k -ésimo tempo;
- e_{ijk} é o erro aleatório correspondente à observação do k -ésimo tempo para a i -ésima dieta no j -ésimo bloco, em que $e \sim N(\phi, \Sigma)$ para $i = 1, \dots, I, j = 1, \dots, J$ e $k = 1, \dots, K_{ij}$.

As curvas de lactação foram obtidas através do ajuste do modelo não-linear. O modelo ajustado foi o de Wood (1967), dado por:

$$y_t = at^b \exp(-ct) + \varepsilon_t.$$

Para verificar a identidade de modelos entre as dietas, foi aplicado o teste de F descrito por Bates & Watts (1988) e Regazzi & Silva (2004). As estimativas dos parâmetros e somas de quadrados foram obtidas utilizando o

PROC NLIN do SAS (2003). As hipóteses consideradas para o teste da identidade de modelos foram as seguintes:

- $H_0^{(1)} : a_1 = a_2 = a$: apenas a produção inicial de leite é a mesma para as duas dietas;
- $H_0^{(2)} : b_1 = b_2 = b$: apenas a taxa crescente de produção até o pico é igual para as duas dietas;
- $H_0^{(3)} : c_1 = c_2 = c$: apenas a taxa de declínio da produção é a mesma para ambas as dietas;
- $H_0^{(4)} : a_1 = a_2 = a \text{ e } b_1 = b_2 = b$: apenas a produção inicial e a taxa crescente de produção até o pico são as mesmas para as duas dietas;
- $H_0^{(5)} : a_1 = a_2 = a \text{ e } c_1 = c_2 = c$: apenas a produção inicial e a taxa de declínio são as mesmas para as duas dietas;
- $H_0^{(6)} : b_1 = b_2 = b \text{ e } c_1 = c_2 = c$: apenas a taxa crescente de produção até o pico e a taxa de declínio de produção são iguais para as duas dietas;
- $H_0^{(7)} : a_1 = a_2 = a \text{ e } b_1 = b_2 = b \text{ e } c_1 = c_2 = c$: apenas a produção inicial, a taxa crescente e a taxa de declínio da produção são iguais para as duas dietas.

4 RESULTADOS E DISCUSSÃO

A Figura 1 representa o gráfico de caixas referente à produção de leite dos animais submetidos às duas dietas. Observando o gráfico, percebe-se que não há grande heterogeneidade de variâncias. As observações destoantes das demais não caracterizaram valores atípicos ou *outliers*, mas, devido ao grande número de observações, foram desconsideradas nas análises. A Tabela 1 apresenta o número de observações por vaca, após serem desconsideradas as observações destoantes.

Os resultados de algumas medidas descritivas para o conjunto de dados analisados são apresentados na Tabela 2. As duas dietas apresentam valores semelhantes.

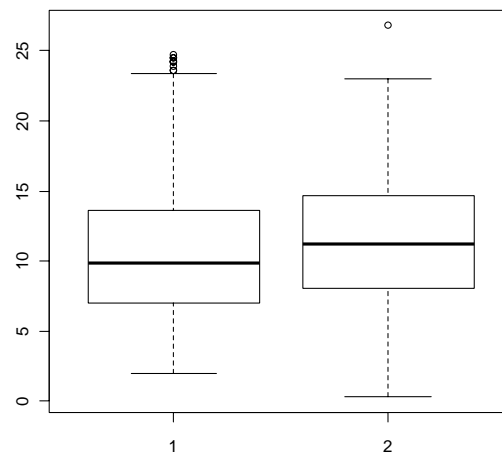


FIGURA 1 Gráfico de caixas da produção de leite (kg) para cada dieta.

TABELA 1 Número de observações/vaca para cada dieta.

Dieta / Vacas	1	2	3	4	5	6	7	8
T ₁	288	263	248	262	266	298	278	268
T ₂	192	305	263	310	295	248	295	235

TABELA 2 Número de observações (n), médias diárias, desvios-padrão (DP) valores mínimos e máximos para a produção de leite, em kg, de cada dieta.

Dieta	n	Médias (kg)	DP (kg)	Mínimo (kg)	Máximo (kg)
T ₁	2171	10,72	4,67	2,00	23,00
T ₂	2143	11,53	4,53	1,70	23,00

4.1 Modelo misto

Para o conjunto de dados, não foi possível modelar a matriz de covariância dos erros (Σ) com nenhuma das estruturas de covariâncias apresentadas. Isto foi devido, talvez, à condição de desbalanceamento dos dados.

Como possível alternativa, selecionou-se dois subconjuntos: uma parte dos dados com um menor grau de desbalanceamento com, no máximo, 269 observações/vaca (caracterizando 50% de lactações desbalanceadas) e, posteriormente, os dados foram analisados na condição balanceada de acordo com o menor tempo de lactação observado, totalizando 192 observações/vaca. Algumas medidas descritivas desses conjuntos de dados estão apresentadas nas Tabelas 3 e 4. As Figuras 2 e 3 representam os gráficos de caixas referentes à produção de leite dos animais submetidos às duas dietas.

TABELA 3 Número de observações (n), médias diárias, desvios-padrão (DP) e valores mínimos e máximos para a produção de leite, em kg, de cada dieta para o conjunto de dados com um menor grau de desbalanceamento.

Dieta	n	Médias (kg)	DP (kg)	Mínimo (kg)	Máximo (kg)
T ₁	2117	10,86	4,65	2,00	23,00
T ₂	2018	11,81	4,43	1,70	23,00

TABELA 4 Número de observações (n), médias diárias, desvios-padrão (DP) e valores mínimos e máximos para a produção de leite, em kg, de cada dieta para o conjunto de dados balanceados.

Dieta	n	Médias (kg)	DP (kg)	Mínimo (kg)	Máximo (kg)
T ₁	1536	12,29	4,54	2,00	23,00
T ₂	1536	13,02	4,41	1,70	23,00

Como pode ser visto nas Tabelas 2, 3 e 4, não foram observadas grandes diferenças entre os valores de desvios-padrão, mínimos e máximos do conjunto de dados completo e dos dois subconjuntos, mas a produção média aumentou no conjunto de dados balanceado. Os gráficos de caixas também são muito semelhantes, não sendo observada grande heterogeneidade de variâncias entre eles.

Para o conjunto de dados com um menor grau de desbalanceamento, bem como para os dados balanceados, foi possível identificar três estruturas de covariâncias razoáveis para a matriz Σ : auto-regressiva de 1^a ordem, componentes de variância e simétrica composta. A Tabela 5 apresenta os resultados obtidos para os índices de AIC e BIC das estruturas de covariâncias modeladas.

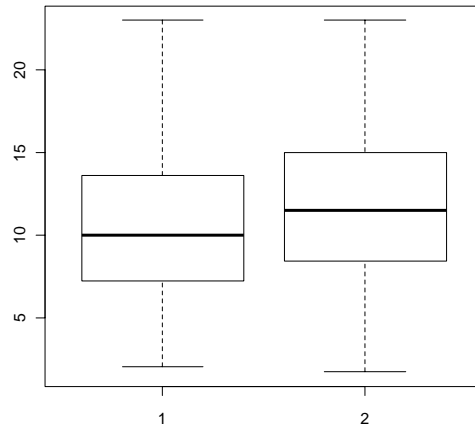


FIGURA 2 Gráfico de caixas da produção de leite (kg) para cada dieta para o conjunto de dados com um menor grau de desbalanceamento.

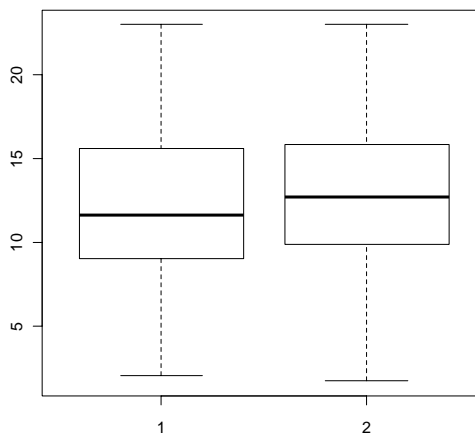


FIGURA 3 Gráficos de caixas para produção de leite (kg) de cada dieta quando o conjunto de dados é balanceado.

TABELA 5 Critérios de informação de Akaike (AIC) e Schwarz (BIC) para os conjuntos de dados com um menor grau de desbalanceamento e balanceados.

Estruturas da Matriz de Covariâncias	Desbalanceados		Balanceados	
	AIC	BIC	AIC	BIC
Auto Regressiva de 1ª Ordem	13372,6	13373,2	10425,9	104426,1
Simétrica Composta	16453,1	16454,7	12469,3	12470,8
Componentes de Variância	17767,4	17768,2	13556,4	13557,2

Conforme os critérios de informação, para ambas as situações, a estrutura da matriz Σ que melhor representa os dados de lactação é a auto-regressiva de 1ª ordem – AR (1), que apresentou menores AIC e BIC do que as estruturas componentes de variância e simétrica composta.

A escolha da estrutura AR (1) para a matriz Σ é coerente com o que ocorre na prática, pois quando a produção de leite é observada, espera-se que haja correlação entre as observações e que essa correlação seja maior entre observações medidas em dias próximos.

As análises de variância para o modelo sob a estrutura AR (1) da matriz Σ para os dois conjuntos de dados estão apresentadas nas Tabelas 6 e 7.

TABELA 6 Análise de variância para os efeitos fixos do modelo com estrutura da matriz Σ do tipo AR (1) para o conjunto de dados com um menor grau de desbalanceamento.

FV	GL	Den. GL*	F	Pr>F
Dietas	1	3583	9,92 **	0,0016
Tempo	268	3583	5,32 **	<0,001
Dieta*Tempo	268	3583	1,00	0,4807

* Grau de liberdade do denominador do teste F.

TABELA 7 Análise de variância para os efeitos fixos do modelo com a estrutura de matriz Σ do tipo AR (1) para o conjunto de dados balanceados.

FV	GL	Den. GL*	F	Pr>F
Dietas	1	2681	6,18 **	0,0130
Tempo	191	2681	4,64 **	<0,001
Dieta*Tempo	191	2681	1,01	0,4453

* Grau de liberdade do denominador do teste F.

Observa-se que os resultados para os dois conjuntos de dados são análogos, havendo significância para os efeitos de Dietas e Tempo. Não foi detectado efeito significativo para a interação entre Dieta e Tempo.

A Figuras 4 representa os modelos não-lineares de Wood (1967) ajustados para as lactações médias por dieta em função dos dias, para os conjuntos: completo com todas as lactações, com um menor grau de desbalanceamento e balanceado. A Tabela 8 apresenta as equações ajustadas para cada conjunto de dados. O coeficiente de determinação (R^2) indicou que as equações se ajustaram bem aos dados de produção de leite, pois a porcentagem de variação na lactação explicada pelos modelos ajustados é de 77,40% para os dados balanceados, 80,42% no caso dos dados com um menor grau de desbalanceamento e 83,16% para os dados completos.

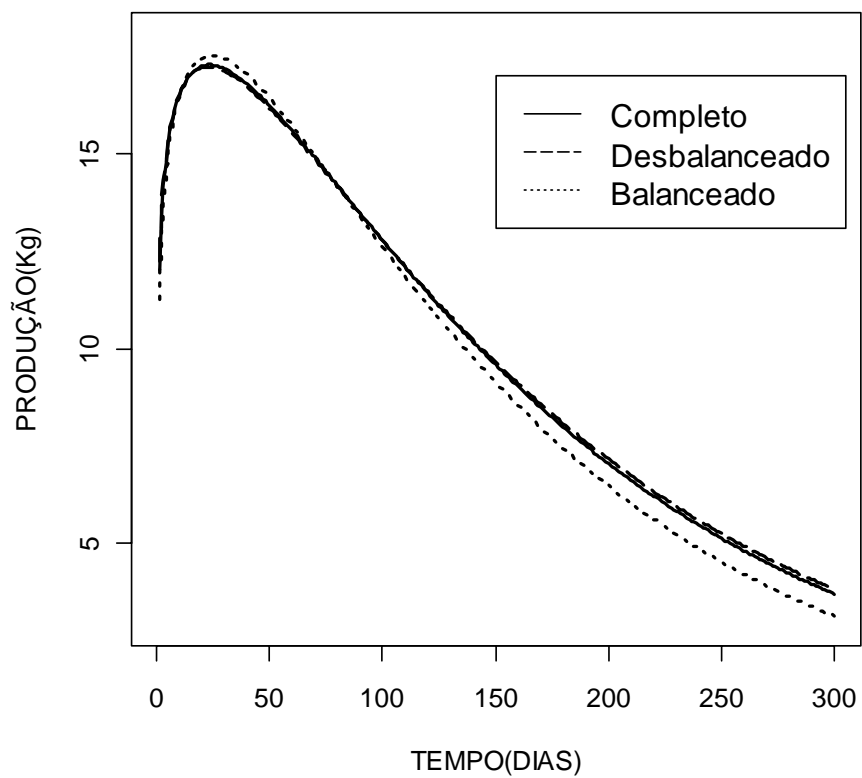


FIGURA 4 Curvas de lactação ajustadas para os conjuntos de dados completo, com um menor grau de desbalanceamento e balanceado.

TABELA 8 Equações ajustadas e coeficientes de determinação (R^2) dos modelos.

Dados	Equações ajustadas	R^2
Completo	$y = 2,4925t^{0,1657} \exp(-0,00708t) + \varepsilon$	83,16%
Desbalanceados	$y = 2,5097t^{0,1577} \exp(-0,00688t) + \varepsilon$	80,42%
Balanceados	$y = 2,4300t^{0,1966} \exp(-0,00801t) + \varepsilon$	77,40%

4.2 Modelo fixo (parcelas subdivididas no tempo)

Não foi possível realizar o teste de esfericidade da matriz Σ devido, possivelmente, ao pequeno número de repetições. Mesmo assim, o ajuste do modelo fixo foi realizado. Como não houve efeito significativo para a interação $(\beta\gamma)_{jk}$, este efeito foi incluído no erro b do modelo fixo apresentado. A Tabela 9 apresenta os resultados da análise univariada no esquema de parcelas subdivididas no tempo para o conjunto completo de dados .

TABELA 9 Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para o conjunto de dados com todas as lactações.

FV	GL	SQ	QM	Fc
Dieta	1	425,5059	425,5059	0,50
Bloco	7	6298,1314	899,7331	1,06
Erro a	7	5954,9889	850,7127	
Tempo	309	60523,3673	195,8685	53,57**
Dieta*Tempo	297	1626,4973	5,4764	1,50**
Erro b	3692	13498,0922	3,6560	

CVparcela = 262,29% CVsubparcela = 17,19%

** (P<0,01).

Não foi observado efeito significativo para os efeitos de Dieta e Tempo, mas verifica-se efeito significativo para a interação entre Dieta e Tempo. Assim, ao contrário do resultado obtido pelo ajuste do modelo misto, o efeito de tempo é diferente nas duas dietas.

O coeficiente de variação da parcela apresenta um valor muito alto, como pode ser observado na Tabela 9. Esse resultado mostra que a variação devido à interação entre Bloco e Dieta (que representa o Erro a) é muito alta, indicando problemas de falta de homogeneidade dentro de blocos. Assim, o modelo fixo não consegue controlar essa variação indevida, que influencia diretamente o teste F, podendo mascarar um possível efeito significativo entre dietas, ao contrário do modelo misto ajustado, que corrige a heterogeneidade dentro de blocos e verifica efeito significativo entre dietas. Além disso, o grau de liberdade do erro b é muito alto, o que pode causar uma sensibilidade do teste F, implicando em uma possível superestimação da interação entre Dieta e Tempo.

O modelo de parcelas subdivididas no tempo também foi ajustado para os conjuntos com um menor grau de desbalanceamento e balanceado como no caso de modelo misto. Os critérios de informação do ajuste do modelo fixo estão apresentados na Tabela 10.

TABELA 10 Critérios de informação para o ajuste do modelo fixo.

Conjunto de dados	AIC	BIC
Balanceados	12379,0	12384,9
Menor grau de desbalanceamento	16052,6	16058,8
Todos os dados	16491,4	16497,7

Comparando os resultados da Tabela 10, obtidos pelo ajuste do modelo fixo, com os valores de AIC e BIC para a estrutura Simetria Composta da Tabela 5, observa-se que em todos os casos (exceto o conjunto de dados completo), o modelo misto ajustado anteriormente apresentou melhor ajuste que o modelo fixo. Possivelmente, esse resultado foi devido ao menor número de parâmetros do modelo misto em relação ao fixo.

As análises de variância no esquema de parcelas subdivididas no tempo para os conjuntos parciais estão apresentadas nas Tabelas 11 e 12. Para ambas as situações, observa-se efeito significativo da interação entre Dieta e Tempo, consistente com a análise para o conjunto com todas as observações. Assim, percebe-se que o grau de desbalanceamento não interferiu no resultado da análise univariada no esquema de parcelas subdivididas no tempo.

TABELA 11 Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para o conjunto de dados com um menor grau de desbalanceamento.

FV	GL	SQ	QM	Fc
Dieta	1	568,4535	568,4535	1,06
Bloco	7	6143,2245	877,7035	0,68
Erro a	7	5820,5411	831,5059	
Tempo	269	56988,5064	211,8532	56,70**
Dieta*Tempo	269	1602,5337	5,9574	1,59**
Erro b	3581	13379,7634	3,7363	

CVparcela = 261,70% CVsubparcela = 17,07%

** (P<0,01).

TABELA 12 Análise de Variância em esquema de parcela subdividida no tempo para produção de leite (kg) para os dados balanceados.

FV	GL	SQ	QM	Fc
Dieta	1	408,9169	408,9196	0.55
Bloco	7	4541,5251	648,7893	0.88
Erro a	7	5167,1123	738,1590	
Tempo	191	36931,3081	193,3576	44,63 **
Dieta*Tempo	191	1389,9169	7,2771	1,68 **
Erro b	2674	11585,4347	4,33	

CVparcela = 201,27% CVsubparcela = 16,45%

** (P<0,01).

4.2.1 Efeito do tempo na produção de cada tratamento

Os modelos de Wood (1967) ajustados para as médias de cada dieta para todos os dados estão apresentados na Tabela 13. Os coeficientes de determinação (R^2) indicaram que os modelos apresentaram ajustes razoáveis para ambas as dietas, pois a porcentagem de variação na lactação explicada pelo modelo ajustado para a dieta 1 é de 70,22%. Para a dieta 2, o coeficiente de variação foi 60,48%. As curvas de lactação ajustadas para as duas dietas estão apresentadas na Figura 6.

TABELA 13 Equações ajustadas e coeficientes de determinação (R^2) do modelo completo.

Dietas	Equações ajustadas	R^2
T ₁	$y = 2,6415t^{0,1062} \exp(-0,00639t) + \varepsilon$	70,22%
T ₂	$y = 2,5294t^{0,1317} \exp(-0,0546t) + \varepsilon$	60,48%

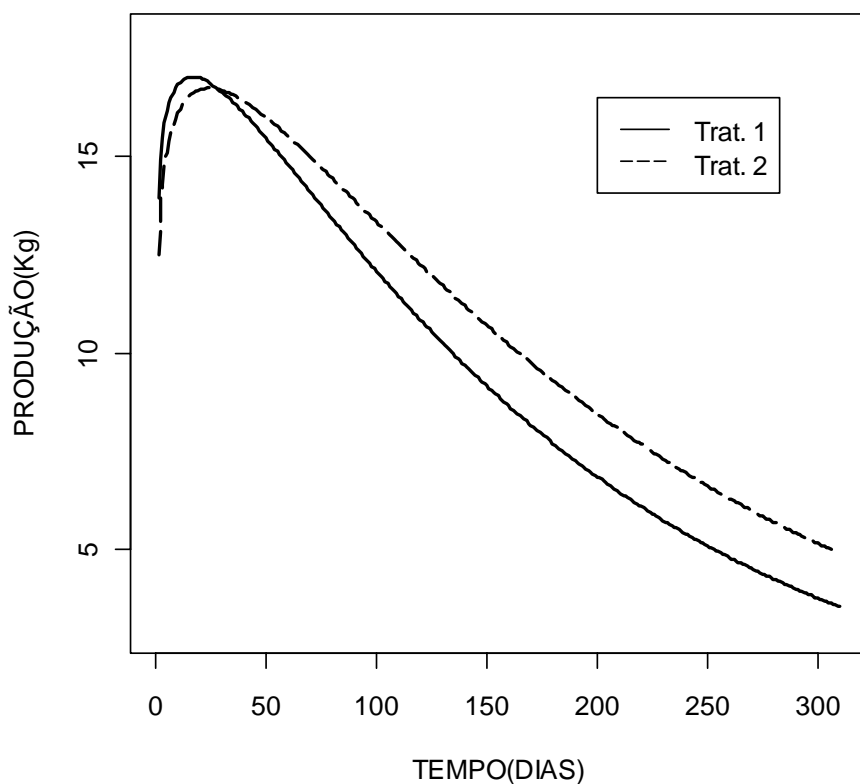


FIGURA 5 Curvas de lactação das duas dietas para o conjunto de dados com todas as lactações.

Foi realizado o teste de identidade de modelos, verificando-se a possibilidade de as equações ajustadas serem representadas por uma equação comum, ou se existe igualdade de pelo menos algum parâmetro nos modelos. As equações estimadas sob as hipóteses de identidade de modelos e igualdade de parâmetros estão apresentadas na Tabela 14. Os resultados dos testes F para as hipóteses formuladas estão apresentados na Tabela 15.

TABELA 14 Equações estimadas para o conjunto de dados completo.

Hipóteses	Equações estimadas
$H_0^{(1)}: a_1=a_2=a, b_1=b_2=b \text{ e } c_1=c_2=c$	$y = 2,5904t^{0,1182} \exp(-0,00582t) + \varepsilon$
$H_0^{(2)}: a_1=a_2=a$	$T_1: y = 2,5873t^{0,1263} \exp(-0,00648t) + \varepsilon$ $T_2: y = 2,5873t^{0,1135} \exp(-0,00525t) + \varepsilon$
$H_0^{(3)}: b_1=b_2=b$	$T_1: y = 2,6101t^{0,1200} \exp(-0,00643t) + \varepsilon$ $T_2: y = 2,5634t^{0,1200} \exp(-0,00531t) + \varepsilon$
$H_0^{(4)}: c_1=c_2=c$	$T_1: y = 2,6909t^{0,0878} \exp(-0,00585t) + \varepsilon$ $T_2: y = 2,4741t^{0,1534} \exp(-0,00585t) + \varepsilon$
$H_0^{(5)}: a_1=a_2=a \text{ e } b_1=b_2=b$	$T_1: y = 2,5887t^{0,1189} \exp(-0,00623t) + \varepsilon$ $T_2: y = 2,5887t^{0,1189} \exp(-0,00545t) + \varepsilon$
$H_0^{(6)}: a_1=a_2=a \text{ e } c_1=c_2=c$	$T_1: y = 2,5896t^{0,1099} \exp(-0,00580t) + \varepsilon$ $T_2: y = 2,5896t^{0,1258} \exp(-0,00580t) + \varepsilon$
$H_0^{(7)}: b_1=b_2=b \text{ e } c_1=c_2=c$	$T_1: y = 2,5619t^{0,1184} \exp(-0,00580t) + \varepsilon$ $T_2: y = 2,6142t^{0,1184} \exp(-0,00580t) + \varepsilon$

TABELA 15 Hipóteses avaliadas e valores da estatística do teste F para o conjunto de dados com todas as lactações.

Hipóteses	F _{calc}
$H_0^1 : a_1 = a_2 = a$	16307,5 **
$H_0^2 : b_1 = b_2 = b$	16327,2 **
$H_0^3 : c_1 = c_2 = c$	16284,0 **
$H_0^4 : a_1 = a_2 = a \quad e \quad b_1 = b_2 = b$	20337,6 **
$H_0^5 : a_1 = a_2 = a \quad e \quad c_1 = c_2 = c$	20078,6 **
$H_0^6 : b_1 = b_2 = b \quad e \quad c_1 = c_2 = c$	19902,4 **
$H_0^7 : a_1 = a_2 = a \quad e \quad b_1 = b_2 = b \quad e \quad c_1 = c_2 = c$	26194,5 **

**($P < 0,01$).

Os resultados dos testes F foram significativos em todas as possíveis situações, considerados os três parâmetros (a , b e c). Isso indica que os modelos ajustados para as duas dietas não são idênticos, não sendo possível representá-los por uma equação comum.

Uma vez que não foi verificada a identidade de modelos, observa-se que o efeito temporal é diferente para as duas dietas. Esse resultado é consistente com a análise no esquema de parcelas subdivididas no tempo, que detectou interação entre dieta e tempo, o que não aconteceu quando do ajuste do modelo misto.

Observando as estimativas dos parâmetros das equações ajustadas para as duas dietas (Tabela 13), pode-se perceber que a dieta 1 apresenta estimativas dos parâmetros a e c superiores às da dieta 2. Apenas a estimativa do parâmetro b é maior para a dieta 2. Isso indica que a produção inicial (a) e a taxa de declínio(c) são maiores quando a suplementação alimentar é suspensa mais cedo

(60 dias após o parto) e que a taxa crescente de produção até o pico (*b*) é superior nos animais que receberam suplementação por mais tempo.

Na prática, o fato de a taxa de declínio da produção ser inferior na dieta 2 pode ser explicado pela condição alimentar diferenciada, como uma consequência de se manter a suplementação alimentar por mais tempo. Essa condição garante que estes animais apresentem maior produção de leite por mais tempo quando comparados àqueles em que a suplementação alimentar foi retirada mais cedo.

O fato de a produção inicial ser maior, quando a suplementação alimentar é suspensa mais cedo -- assim como a taxa crescente de produção ser superior, quando a suplementação é mantida por mais tempo -- sugere que os animais dentro de blocos não são homogêneos. De acordo com a interpretação biológica desses parâmetros, estes representam situações nas quais todos os animais, independente da dieta, estavam recebendo a mesma suplementação alimentar. Esse resultado confirma as afirmações a respeito do ajuste do modelo fixo, indicando que este não foi apropriado para esse conjunto de dados de lactação, pois não corrige a heterogeneidade dentro de bloco.

Para comparação das dietas, foi ajustado o modelo de Wood (1967) para os conjuntos com um menor grau de desbalanceamento e balanceado, obtendo-se resultados análogos aos apresentados pelo conjunto de dados completo. Os modelos ajustados para cada dieta, as curvas de lactação, as equações estimadas para sob as hipóteses de identidade de modelos e igualdade de parâmetros e os testes F para as hipóteses formuladas estão apresentados em anexo.

5 CONCLUSÕES

- O ajuste do modelo fixo (parcela subdividida no tempo) não foi apropriado para esse conjunto de dados de lactação;
- O modelo misto ajustado foi adequado a este conjunto de dados de lactação;
- A estrutura da matriz de covariâncias dos erros mais apropriada para esse conjunto de dados foi a auto-regressiva de 1ª ordem;
- O modelo não-linear ajustado reforça a escolha do modelo misto como apropriado a esse conjunto de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, Cambridge, v. 60, n. 2, p. 255-265, Aug. 1973.
- AKAIKE, H. On entropy maximization principle. In: KRISHNAIAH, P. R. (Ed.). **Applications of statistics**. Amsterdam: North-Holland, 1977. p. 27-41.
- ALI, T. E.; SCHAEFFER, L. R. Accounting for covariances among test day milk yields in dairy cows. **Canadian Journal of Animal Science**, Ottawa, v. 67, n. 3, p. 637-644, Sept. 1987.
- ARAÚJO, C. V. de; TORRES, R. A. de; COSTA, C. N.; TORRES FILHO, R. A. de; ARAÚJO, S. I.; LOPES, P. S.; REGAZZI, A. J.; PEREIRA, C. S.; SARMENTO, J. L. R. Uso de modelos de regressão aleatória para descrever a variação genética da produção de leite da raça holandesa. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 35, n. 3, p. 975-981, 2006. Suplemento.
- ARAÚJO, M. S. A. **Modelos não lineares para dados de contagem longitudinais**. 2007. 198 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: J. Wiley, 1988. 365 p.
- BOX, G. E. P. Some theorems on quadratics applied to the study of analysis of variance problems. **Annals of Mathematical Statistics**, Ann Arbor, v. 25, n. 2, p. 290-302, 1954.
- BRODY, S.; RAGSDALE, A. C.; TURNER, C. W. The rate of decline of milk secretion with the advance of the period of lactation. **The Journal of General Physiology**, New York, v. 5, n. 4, p. 441-444, Mar. 1923.
- COBUCCI, J. A.; EUCLYDES, R. F.; VERNEQUE, R. S.; TEODORO, R. L.; LOPES, P. S. de; SILVA, M. de A. e. Curva de lactação na raça Guzerá. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 29, n. 5, p. 1332-1339, set./out. 2000.

COSTA, S. C. da. **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.

COSTA, S. C. da; DEMÉTRIO, C. G. B.; SUGUINO, E.; GIOLO, S. R. Modelando dados longitudinais de proporções de pegamento de camu-camu. **Revista de Matemática e Estatística**, São Paulo, v. 22, n. 3, p. 113-127, dez. 2004.

DORNELES, C. K. P. **Estudo da curva e persistência da lactação de vacas da raça holandesa utilizando modelo de regressão aleatória**. 2006. 110 p. Dissertação (Mestrado em Melhoramento Genético Animal) – Universidade Federal de Santa Maria, Santa Maria.

EL FARO, L.; ALBUQUERQUE, L. G. Utilização de modelos de regressão aleatória para produção de leite no dia de controle, com diferentes estruturas de variâncias residuais. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 32, n. 5, p. 1104-1113, set./out. 2003.

FERREIRA, E. B. Revisão sobre a utilização da regressão aleatória: um exemplo no ajuste de curvas de lactação de gado bovino. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 48.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 10., 2003, Lavras. **Anais...** Lavras: SEAGRO, 2003. 1 CD-ROM.

FREITAS, E. G. **Análise de dados longitudinais em experimentos com cana-de-açúcar**. 2007. 74 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

GEISSER, S.; GREENHOUSE, S. W. An extension of Box's results on the use of F distribution in multivariate analysis. **Annals of Mathematical Statistics**, Ann Arbor, v. 29, n. 3, p. 855-891, 1958.

GEISSER, S.; GREENHOUSE, S. W. On methods in the analysis of profile data. **Psychometrika**, Williamsburg, v. 24, n. 2, p. 95-112, June 1959.

GILL, J. L. **Design and analysis of experiments in the animal and medical sciences**. Ames: The Iowa State University, 1986. 310 p.

GONÇALVES, T. M.; OLIVEIRA, A. I. G. de; FREITAS, R. T. F. de; PEREIRA, I. G. C. Lactação em rebanhos da raça holandesa no Estado de Minas Gerais: escolha do modelo de melhor ajuste. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 31, n. 4, p. 1689-1694, jul./ago. 2002.

GRAYBILL, F. A. **Theory and application of the linear model**. Belmont: Duxbury, 1976. 70 p.

GROENEWALD, P. C. N.; VILJOEN, C. S. A Bayesian model for analysis of lactation curves of dairy goats. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 8, n. 1, p. 75-83, Mar. 2003.

HENDERSON JÚNIOR, C. R. Analysis of covariance in the mixed model: higher-level, nonhomogeneous and random regression. **Biometrics**, Cambridge, v. 38, n. 3, p. 623-640, Sept. 1982.

HUYNH, H.; FELDT, L. Condition under which mean square ratios in repeated measurements designs have exact F-distributions. **Journal American Statistics Association**, Alexandria, v. 65, n. 332, p. 1582-1589, Dec. 1970.

HUYNH, H.; FELDT, L. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. **Journal of Education Association**, Ontario, v. 1, n. 1, p. 69-82, 1976.

LIANG, K. Y.; ZEGER, S. L. Longitudinal data analysis using generalized linear models. **Biometrika**, Cambridge, v. 73, n. 1, p. 13-22, Feb. 1986.

LIMA, C. G. **Análises de dados longitudinais provenientes de experimentos em blocos casualizados**. 1996. 126 p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

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

MALHEIROS, E. B. Precisão da análise de dados longitudinais com diferentes estruturas para matriz de variância e covariância quando se utiliza esquema de parcelas subdivididas. **Revista de Matemática e Estatística**, São Paulo, v. 1, n. 17, p. 1229-1239, abr. 1999.

MAUCHLY, J. W. Significance test for sphericity of a normal n-variate distribution. **Annals of Mathematical Statistics**, Ann Arbor, v. 11, n. 2, p. 204-209, 1940.

MORRISON, D. F. **Multivariate statistical methods**. 3. ed. New York: McGraw-Hill, 1990. 415 p.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal Royal Statistical Society Series A**, London, v. 135, n. 3, p. 370-384, May 1972.

NETER, J.; WASSERMAN, W. **Applied linear statistical models: regression, analysis of variance and experimental design**. Homewood: R.D. Irwin, 1990. 842 p.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. **Revista Ceres**, Viçosa, MG, v. 40, n. 228, p. 176-195, 1993.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de parâmetros no caso de dados de delineamentos experimentais. **Revista Ceres**, Viçosa, MG, v. 46, n. 266, p. 383-409, 1999.

REGAZZI, A. J.; SILVA, C. H. O. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear: I., dados de experimentos. **Revista de Matemática e Estatística**, São Paulo, v. 22, n. 3, p. 33-45, jul. 2004.

RIBOLDI, J. Modelos mistos aplicados. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 52.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12., 2007, Santa Maria. **Anais...** Santa Maria: SEAGRO, 2007. 1 CD-ROM.

ROCHA, F. M. M. **Seleção de estruturas de covariância para dados com medidas repetidas**. 2004. 114 p. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.

ROSÁRIO, M. F.; SILVA, M. A. N.; SAVINO, V. J. M.; COELHO, A. A. D.; MORAES, M. C. Avaliação do desempenho zootécnico de genótipos de frangos de corte utilizando-se a análise de medidas repetidas. **Revista Brasileira de Zootecnia**, Viçosa, MG, v. 34, n. 6, p. 2253-2261, 2005. Suplemento.

STATISTICAL ANALYSIS SYSTEM INSTITUTE. **SAS/STAT software: changes and enhancements through release 9.1.** Cary, 2003. 5136 p.

SHERCHAND, L.; McNEW, R. W.; KELLOGG, D. W. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. **Journal Dairy Science**, Champaign, v. 78, n. 11, p. 2507-2513, Nov. 1995.

SILVA, F. F. **Abordagem Bayesiana da curva de lactação de cabras Saanen.** 2003. 64 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics: a biometrical approach.** 2. ed. New York: MacGraw-Hill, 1980. 633 p.

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, Universidade de São Paulo, São Paulo.

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

WOOD, P. D. P. Algebraic model of the lactation curve in cattle. **Nature**, London, v. 216, n. 25, p. 164-165, Oct. 1967.

ANEXOS

ANEXO A	Páginas
TABELA 1A	Equações ajustadas e coeficientes de determinação (R^2) do modelo completo para os conjuntos de dados com um menor grau de desbalanceamento e balanceado42
TABELA 2A	Equações estimadas para o conjunto de dados com um menor grau de desbalanceamento.....43
TABELA 3A	Equações estimadas para o conjunto de dados balanceados.....44
TABELA 4A	Hipóteses avaliadas e valores da estatística do teste F para os conjuntos de dados com um menor grau de desbalanceamento e balanceado.....45

TABELA 1A Equações ajustadas e coeficientes de determinação (R^2) do modelo completo para os conjuntos de dados com um menor grau de desbalanceamento e balanceado.

Dados	Equações ajustadas	R^2
Desbalanceados	$y_1 = 2,6321t^{0,1108} \exp(-0,00651t) + \varepsilon$	69,34%
	$y_2 = 2,5119t^{0,1452} \exp(-0,00581t) + \varepsilon$	61,48%
Balanceados	$y_1 = 2,4707t^{0,1928} \exp(-0,00899t) + \varepsilon$	68,24%
	$y_2 = 2,3826t^{0,2061} \exp(-0,00745t) + \varepsilon$	55,45%

TABELA 2A Equações estimadas para o conjunto de dados com um menor grau de desbalançamento.

Hipóteses	Equações estimadas
$H_0^{(1)}$: $a_1=a_2=a$, $b_1=b_2=b$ e $c_1=c_2=c$	$y = 2,5668t^{0,1293} \exp(-0,00609t) + \varepsilon$
$H_0^{(2)}$: $a_1=a_2=a$	$T_1: y = 2,5669t^{0,1340} \exp(-0,00662t) + \varepsilon$ $T_2: y = 2,5669t^{0,1247} \exp(-0,00557t) + \varepsilon$
$H_0^{(3)}$: $b_1=b_2=b$	$T_1: y = 2,5859t^{0,1296} \exp(-0,00661t) + \varepsilon$ $T_2: y = 2,5470t^{0,1296} \exp(-0,00560t) + \varepsilon$
$H_0^{(4)}$: $c_1=c_2=c$	$T_1: y = 2,6634t^{0,0999} \exp(-0,00611t) + \varepsilon$ $T_2: y = 2,4583t^{0,1619} \exp(-0,00611t) + \varepsilon$
$H_0^{(5)}$: $a_1=a_2=a$ e $b_1=b_2=b$	$T_1: y = 2,5673t^{0,1290} \exp(-0,00644) + \varepsilon$ $T_2: y = 2,5673t^{0,1290} \exp(-0,00573t) + \varepsilon$
$H_0^{(6)}$: $a_1=a_2=a$ e $c_1=c_2=c$	$T_1: y = 2,5668t^{0,1213} \exp(-0,00607t) + \varepsilon$ $T_2: y = 2,5668t^{0,1360} \exp(-0,00607t) + \varepsilon$
$H_0^{(7)}$: $b_1=b_2=b$ e $c_1=c_2=c$	$T_1: y = 2,5412t^{0,1293} \exp(-0,00607t) + \varepsilon$ $T_2: y = 2,5891 t^{0,1293} \exp(-0,00607t) + \varepsilon$

TABELA 3A Equações estimadas para o conjunto de dados balanceados.

Hipóteses	Equações estimadas
$H_0^{(1)}$: $a_1=a_2=a$, $b_1=b_2=b$ e $c_1=c_2=c$	$y = 2,4397t^{0,1943} \exp(-0,00785t) + \varepsilon$
$H_0^{(2)}$: $a_1=a_2=a$	$T_1: y = 2,4224t^{0,2112} \exp(-0,00907t) + \varepsilon$ $T_2: y = 2,4224t^{0,1885} \exp(-0,00716t) + \varepsilon$
$H_0^{(3)}$: $b_1=b_2=b$	$T_1: y = 2,4563t^{0,2002} \exp(-0,00891t) + \varepsilon$ $T_2: y = 2,3862t^{0,2002} \exp(-0,00732t) + \varepsilon$
$H_0^{(4)}$: $c_1=c_2=c$	$T_1: y = 2,5398t^{0,1644} \exp(-0,00809t) + \varepsilon$ $T_2: y = 2,2899t^{0,2385} \exp(-0,00809t) + \varepsilon$
$H_0^{(5)}$: $a_1=a_2=a$ e $b_1=b_2=b$	$T_1: y = 2,4254t^{0,1980} \exp(-0,00850t) + \varepsilon$ $T_2: y = 2,4254t^{0,1980} \exp(-0,00760t) + \varepsilon$
$H_0^{(6)}$: $a_1=a_2=a$ e $c_1=c_2=c$	$T_1: y = 2,4267t^{0,1899} \exp(-0,00800t) + \varepsilon$ $T_2: y = 2,4267t^{0,2039} \exp(-0,00800t) + \varepsilon$
$H_0^{(7)}$: $b_1=b_2=b$ e $c_1=c_2=c$	$T_1: y = 2,40501t^{0,1973} \exp(-0,00801t) + \varepsilon$ $T_2: y = 2,4465 t^{0,1973} \exp(-0,00801t) + \varepsilon$

TABELA 4A Hipóteses avaliadas e valores da estatística do teste F para os conjuntos de dados com um menor grau de desbalanceamento e balanceado.

Hipóteses	F _{calc}	
	Desbalanceados	Balanceados
$H_0^1 : a_1 = a_2 = a$	15973,5 **	13899,7 **
$H_0^2 : b_1 = b_2 = b$	15994,4 **	13920,8 **
$H_0^3 : c_1 = c_2 = c$	16005,3 **	13851,1 **
$H_0^4 : a_1 = a_2 = a \quad e \quad b_1 = b_2 = b$	19945,5 **	17250,3 **
$H_0^5 : a_1 = a_2 = a \quad e \quad c_1 = c_2 = c$	19671,5 **	16985,9 **
$H_0^6 : b_1 = b_2 = b \quad e \quad c_1 = c_2 = c$	19609,8 **	16843,5 **
$H_0^7 : a_1 = a_2 = a \quad e \quad b_1 = b_2 = b \quad e \quad c_1 = c_2 = c$	25856,7 **	22246,0 **

**($P < 0,01$).

ANEXO B

Páginas

FIGURA 1B	Curvas de lactação das duas dietas para o conjunto de dados com um menor grau de desbalanceamento.....	46
FIGURA 2B	Curvas de lactação das duas dietas para o conjunto de dados balanceados.....	46

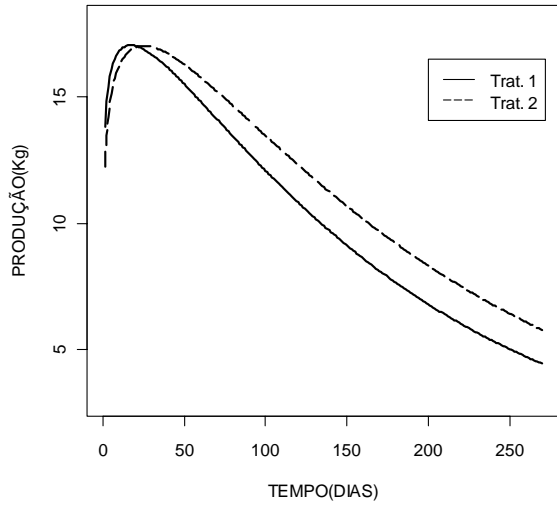


FIGURA 1B. Curvas de lactação das duas dietas para o conjunto de dados com um menor grau de desbalanceamento.

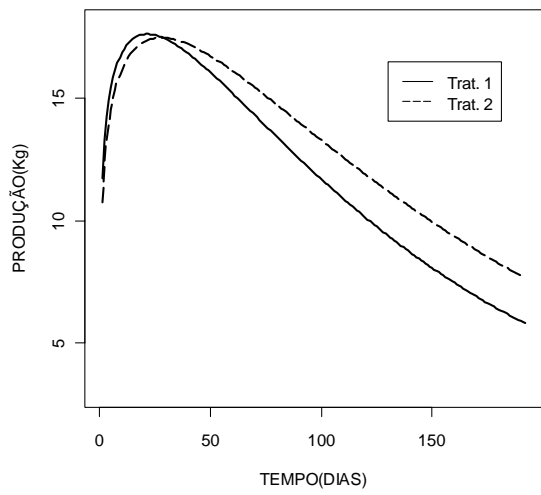


FIGURA 2B. Curvas de lactação das duas dietas para o conjunto de dados balanceados.