



**WEDERSON LEANDRO FERREIRA**

**ANÁLISE DE DADOS COM MEDIDAS  
REPETIDAS EM EXPERIMENTO COM  
INGESTÃO DE CAFÉ**

**LAVRAS - MG**

**2012**

**WEDERSON LEANDRO FERREIRA**

**ANÁLISE DE DADOS COM MEDIDAS REPETIDAS EM  
EXPERIMENTO COM INGESTÃO DE CAFÉ**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador

Dr. Augusto Ramalho de Moraes

**LAVRAS - MG**

**2012**

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

Ferreira, Wederson Leandro.

Análise de dados com medidas repetidas em experimento com ingestão de café / Wederson Leandro Ferreira. – Lavras : UFLA, 2012.

108 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2012.

Orientador: Augusto Ramalho de Moraes.

Bibliografia.

1. Dados longitudinais. 2. Modelo linear misto. 3. Análise de perfil. 4. Critério de informação. I. Universidade Federal de Lavras. II. Título.

CDD – 519.53

**WEDERSON LEANDRO FERREIRA**

**ANÁLISE DE DADOS COM MEDIDAS REPETIDAS EM  
EXPERIMENTO COM INGESTÃO DE CAFÉ**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 8 de fevereiro de 2012.

Dr. Adriano Ferreti Borgatto                      UFSC

Dr. José Airton Rodrigues Nunes              UFLA

Dr. João Domingos Scalon                      UFLA

Dr. Augusto Ramalho de Moraes

Orientador

**LAVRAS - MG**

**2012**

*A minha família, em especial Maria, minha mãe, que jamais mediu esforços em valorizar todas as etapas educacionais que alcancei até o momento.*

*A todos os amigos que deram, e ainda me dão, exemplos de caráter, honestidade e bom senso, contribuindo, dessa forma, para as vitórias até aqui alcançadas.*

## **DEDICO**

## AGRADECIMENTOS

A Deus, agradeço por todas as coisas boas que vivi, porque sei que o bem apenas dele é que veio.

A minha mãe, pelo exemplo de sempre.

A meus irmãos e sobrinhos, pelo apoio que sempre me deram, cada um a seu modo.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade de poder cursar o mestrado, período em que muito fui feliz e aprendi.

À CAPES, pela concessão da bolsa de estudos.

Ao professor Dr. Augusto Ramalho de Moraes, pela orientação, pelos ensinamentos e dedicação, e com quem muito aprendi.

Ao CNPq, pela concessão de auxílio financeiro para o projeto de pesquisa.

Aos demais professores e funcionários do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária que tiveram papel fundamental em meu crescimento acadêmico.

Aos amigos do mestrado que, a partir de agora, fazem parte da minha família.

## RESUMO

Este trabalho se propôs a abordar técnicas de análise de medidas repetidas no tempo, com ênfase prioritária no ajuste de curvas polinomiais de crescimento para o peso de animais (ratos) submetidos a diferentes dietas alimentares, com e sem extrato aquoso de café da espécie *Coffea arabica*. Os animais foram avaliados durante 15 semanas. Para o estudo dos pesos dos animais, foram utilizadas diferentes técnicas estatísticas. Análise exploratória, análise univariada, utilizando o delineamento inteiramente casualizado em esquema de parcelas subdivididas no tempo, que impõe forte restrição quanto à matriz de covariâncias do erro experimental, análise multivariada que permite o uso de uma matriz de covariâncias não estruturada e modelo linear misto, que permite selecionar as estruturas de covariâncias que melhor se adequam ao perfil de correlação dos dados. No ajuste de curvas de crescimento, usando modelo linear misto, utilizaram-se diferentes estruturas de covariâncias para descrição da correlação presente em uma mesma unidade experimental. Em todas as técnicas utilizadas, houve resultados significativos para os tratamentos. Procedendo-se ao ajuste das curvas polinomiais de crescimento do peso médio dos animais para as diferentes dietas ao longo das semanas, constatou-se que a estrutura de covariâncias UN (não estruturada) relacionada aos parâmetros de efeitos aleatórios e a estrutura VC (componente de variância) relacionada ao erro experimental, que possui variâncias iguais e correlações nulas, foram as mais adequadas, por ambos os critérios de informação, Akaike e BIC. Somente na dieta (NORM), que recebeu ração de biotério enriquecida com ração para cães, o uso do café apresentou tendência a diminuir o peso dos animais, ao longo das semanas avaliadas.

Palavras-chave: Dados longitudinais. Modelo linear misto. Análise de perfil. Critério de informação.

## ABSTRACT

This study intended to approach the analysis of repeated measures technique, with an emphasis on the fitting of polynomial growth curves for the weight of animals (rats) submitted to diets with or without aqueous extract of Arabian coffee (*Coffea Arabica*). The animals were evaluated for 15 weeks. To study the weights of the animals, we used different statistical techniques. Exploratory analysis, univariate analysis utilizing a randomized design with a split-plot on time arrangement, which constraints the covariance matrix of the experimental error; multivariate analysis, which enables the use of a non-structured covariance matrix and a linear mixed model, which enables the selection of covariance structures that best fit the data correlation profile. In the setting of growth curves, using linear mixed model, we used different covariance structures for describing the correlation present in the same experimental unit. In all techniques, there were significant results for the treatments. Following the fitting of polynomial growth curves for the average weight of animals under different diets over the weeks, it was found that the unstructured (UN) covariance structure related to the random effect parameters as well as the VC structure (variance component) related to experimental error, which has equal variances and null correlations, were the most appropriate to both information criteria; Akaike and BIC. Only under the diet (NORM), which received bioterio ration enriched with dog ration, the use of coffee tended to decrease the weight of the animals along the evaluated weeks.

Keywords: Longitudinal data. Linear mixed model. Profile analysis. Information criterion.

## LISTA DE FIGURAS

Figura 1	Representação gráfica do perfil individual do peso dos animais, em gramas, em função dos tempos de avaliação (semanas), para as diferentes dietas alimentares.....	63
Figura 2	Representação gráfica do peso dos animais, em gramas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio em cada unidade experimental.....	64
Figura 3	Representação gráfica do peso dos animais, em gramas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio do peso dos animais em cada semana avaliada.....	65
Figura 4	Representação gráfica do peso dos animais, em gramas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio em cada dieta alimentar.....	66
Figura 5	Representação gráfica do peso médio dos animais em relação ao período de avaliação (semanas), para as seis dietas utilizadas.....	67
Figura 6	Representação gráfica da probabilidade normal dos resíduos das curvas ajustadas ao peso, em gramas, das dietas alimentares, na evolução das 15 semanas avaliadas.....	81
Figura 7	Descrição gráfica do crescimento de ratos por meio de curvas polinomiais ajustadas ao peso por animal (em gramas), durante período de 15 semanas, para cada dieta alimentar.....	86

## LISTA DE TABELAS

Tabela 1	Esquema para análise da variância, esperanças dos quadrados médios e teste F para as hipóteses de interesse do modelo de experimento no delineamento inteiramente casualizado em esquema de parcelas subdividas no tempo.....	23
Tabela 2	Resultado da análise de variância univariada da variável peso por animal (grama), considerando o delineamento inteiramente casualizado e com o fator grupos de dietas nas parcelas e os períodos de avaliação nas subparcelas.....	68
Tabela 3	Resultado do teste de esfericidade de Mauchly .....	69
Tabela 4	Resultado da análise univariada de perfil corrigida em um delineamento inteiramente casualizado para os efeitos intraindivíduos do peso do animal submetido a 6 dietas, aferido no período de 15 semanas, a partir da 4 <sup>a</sup> semana de vida .....	70
Tabela 5	Resultado da análise multivariada da variável peso (grama) por animal, considerando o delineamento inteiramente casualizado e testando o efeito nulo do fator semanas.....	73
Tabela 6	Resultado da análise multivariada da variável peso (grama) por animal, considerando o delineamento inteiramente casualizado e testando efeito nulo da interação de semanas × dietas.....	74
Tabela 7	Estimativas, erros padrões, valor do teste <i>t</i> , nível de significância (valor-p) para os parâmetros dos modelos das curvas polinomiais ajustadas para pesos dos animais (em gramas), no período de 15 semanas e coeficiente de determinação para as seis dietas .....	77

Tabela 8	Estimativas dos critérios de seleção das estruturas da matriz de covariâncias para G, variação entre indivíduos e R, variação intra indivíduo, considerando o modelo maximal e o método de estimação da máxima verossimilhança restrita (MVR) .....	80
Tabela 9	Teste para os efeitos fixos para o ajuste do modelo selecionado dos pesos dos animais, em gramas, aferidos no período de 15 semanas, a partir da 4 <sup>a</sup> semana de vida, para as 6 dietas.....	82
Tabela 10	Estimativas e avaliação de contrastes formados por dietas que receberam e não receberam extrato aquoso de café, utilizando a estatística <i>t</i> .....	83
Tabela 11	Estimativas e teste para os parâmetros do modelo das curvas polinomiais ajustadas pelo modelo linear misto para pesos dos animais, em gramas, aferidos no período de 15 semanas, a partir da 4 <sup>a</sup> semana de vida, para as 6 dietas alimentares .....	84

## SUMÁRIO

1	INTRODUÇÃO .....	13
2	REFERENCIAL TEÓRICO .....	15
2.1	Evolução histórica do Café .....	15
2.1.1	Café na nutrição funcional .....	16
2.2	Análise de medidas repetidas no tempo .....	17
2.3	Análise de perfil .....	21
2.3.1	Análise univariada de perfil .....	21
2.3.1.1	Teste de esfericidade de Mauchly .....	27
2.3.1.2	Correções do número de graus de liberdade.....	29
2.3.2	Análise multivariada de perfil.....	31
2.4	Modelo linear misto, com referência a curvas de crescimento .....	39
2.4.1	Estimação do parâmetro $\beta$ e predição de $u$ .....	43
2.4.2	Estimação dos parâmetros das matrizes de covariância $G$ e $R$ .....	45
2.4.3	Estruturas das matrizes de covariâncias.....	47
2.4.4	Seleção dos modelos.....	50
2.4.4.1	Teste Assintótico da Razão de Verossimilhança.....	51
2.4.4.2	CrITÉRIOS de informação.....	51
2.4.5	Inferência sobre os efeitos fixos e aleatórios.....	53
2.5	Análise de medidas repetidas com ingestão de café .....	54
3	MATERIAL E MÉTODOS .....	55
3.1	Material .....	55
3.1.1	Modelo experimental .....	56
3.2	Métodos. ....	56
4	RESULTADOS E DISCUSSÃO .....	63
4.1	Análise exploratória .....	63
4.2	Análise univariada de perfil. ....	67
4.2.1	Teste de Mauchly .....	68
4.2.2	Análise univariada de perfil corrigida para as fontes de variação semanas e interação de dietas com semanas .....	69
4.3	Análise multivariada de perfil.....	71
4.4	Análise por meio do modelo linear misto. ....	76
4.4.1	Escolha dos efeitos fixos associados aos perfis médios de resposta da variável peso.....	76
4.4.2	Identificação dos efeitos aleatórios associados aos perfis individuais de resposta da variável peso.....	78

<b>4.4.3</b>	<b>Ajuste das matrizes de covariâncias, testes, estimação dos parâmetros e diagnóstico do modelo.....</b>	<b>79</b>
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>87</b>
<b>6</b>	<b>TRABALHOS FUTUROS .....</b>	<b>88</b>
	<b>REFERÊNCIAS .....</b>	<b>89</b>
	<b>ANEXOS .....</b>	<b>96</b>

## 1 INTRODUÇÃO

Muitos são os campos da pesquisa nos quais são realizadas várias observações sobre a mesma unidade experimental, como ocorre, por exemplo, na avaliação do crescimento do peso de espécies animais. Planejamentos desse tipo permitem que se avaliem as mudanças que ocorrem ao longo do tempo.

Como as medidas são realizadas nas mesmas unidades experimentais e, em geral, de modo sistemático, é de se esperar que ocorra correlação não nula entre as medidas no tempo, bem como, exista certa heterogeneidade de variâncias. Nesse caso, é possível considerar que as respostas de tempos mais próximos sejam mais correlacionadas do que aquelas de tempos mais distantes, tornando, quase sempre, característica comum a dados mensurados ao longo do tempo (LITTELL et al., 2006).

Tendo em vista este aspecto, é primordial uma abordagem apropriada à análise estatística dos dados que leve em consideração essa possível correlação presente entre as observações em uma mesma unidade experimental porque, caso contrário, todas as inferências realizadas podem ser inválidas ou distorcidas da realidade.

A análise estatística de dados com essa característica pode ser feita utilizando-se técnicas uni ou multivariadas que, geralmente, são direcionadas a dados obtidos nos mesmos instantes de tempo para todos os tratamentos e também com ausência de parcelas perdidas (BARBOSA, 2009). Outro procedimento estatístico que pode ser adotado é a construção de curvas de crescimento, utilizando, por exemplo, modelos lineares mistos.

Ao analisar dados por meio de um modelo univariado utilizando-se o esquema de parcela subdividida no tempo, deve-se analisar a estrutura da matriz de variâncias e covariâncias, pois há a pressuposição de homogeneidade de variâncias e correlações nulas quanto ao uso desta estrutura. Este fato contraria a

correlação serial geralmente existente entre as observações obtidas longitudinalmente de uma mesma unidade experimental. Uma técnica alternativa é a análise multivariada de perfil que não impõe restrição quanto à estrutura da matriz de covariância, porém, tem a desvantagem de não considerar, na análise, dados de unidades experimentais com presença de parcelas perdidas.

O uso de modelos lineares mistos torna-se um procedimento estatístico extremamente versátil quando o intuito é a modelagem de curvas polinomiais de crescimento, pois é possível optar por uma estrutura de covariância que acomode a correlação possivelmente existente nas unidades experimentais.

Neste trabalho, propôs-se abordar as técnicas clássicas uni e multivariada de perfil, ao analisar medidas repetidas no tempo e o ajuste de curvas polinomiais de crescimento por meio do modelo linear misto aplicado em experimento no qual se avaliou o ganho de peso dos animais (ratos), ao longo de 15 semanas, em função de diferentes dietas com e sem ingestão de café.

## **2 REFERENCIAL TEÓRICO**

Nesta seção são apresentados conceitos e métodos que foram utilizados para atingir o objetivo deste trabalho.

### **2.1 Evolução histórica do café**

Não há evidência real sobre a descoberta do café, mas há muitas lendas que relatam sua possível origem. Uma das mais aceitas e divulgadas é a do pastor Kaldi, que viveu na Absínia, hoje Etiópia, há cerca de mil anos. Ela conta que Kaldi, observando suas cabras, notou que elas ficavam alegres e saltitantes e que esta energia extra se evidenciava sempre que mastigavam os frutos de coloração amarelo-avermelhada dos arbustos existentes em alguns campos de pastoreio. A planta de café é originária da Etiópia, onde, ainda hoje, faz parte da vegetação natural. Foi a Arábia a responsável pela propagação da cultura do café. O nome café não é originário da Kaffa, local de origem da planta, e sim da palavra árabe *qahwa*, que significa vinho. Por esse motivo, o café era conhecido como "vinho da Arábia", quando chegou à Europa, no século XIV (ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE CAFÉ - ABIC, 2011).

Ainda segundo a ABIC (2011), o hábito de tomar café foi desenvolvido na cultura árabe. No início, era conhecido apenas por suas propriedades estimulantes e a fruta era consumida fresca, sendo utilizada para alimentar e estimular os rebanhos durante viagens. Com o tempo, o café começou a ser macerado e misturado com gordura animal, para facilitar o consumo durante as viagens. Em 1000 d.C., os árabes começaram a preparar uma infusão com as cerejas, fervendo-as em água. Entretanto, foi no século XIV que o processo de torrefação foi desenvolvido e, finalmente, a bebida adquiriu um aspecto mais parecido com o dos dias de hoje.

Devido à sua grande aceitação na Europa, inicialmente em Veneza, os colonizadores europeus foram responsáveis pela vinda do café ao Suriname, São Domingos, Cuba, Porto Rico e Guianas, chegando, logo em seguida, ao norte do Brasil.

No Brasil, devido às condições climáticas, o cultivo de café se espalhou rapidamente, passando pelo Maranhão, Bahia, Rio de Janeiro, São Paulo, Paraná e Minas Gerais. Num espaço de tempo relativamente curto, o café passou de uma posição relativamente secundária para a de produto-base da economia, tornando o Brasil o maior produtor atualmente, respondendo por cerca de 30% do mercado internacional (ABIC, 2011). Como mercado consumidor, perde apenas para os Estados Unidos. Seu mercado possui previsão de movimentar em 2011 cerca de R\$ 7,5 bilhões, somente no mercado nacional (ABIC, 2011). No mundo, esse mercado perde apenas para o de petróleo.

Algumas pesquisas indicam que o consumo médio de café está aumentando, fato muito justificado pelo seu fator estimulante e seu sabor forte e próprio.

Cecon et al. (2008) relatam que, atualmente, o café é produzido em mais de 60 países, predominantemente em países menos desenvolvidos, sendo atividade de grande aspecto socioeconômico no mundo. E é consumido em países mais ricos, algo que faz com que a sua qualidade se aprimore cada vez mais.

### **2.1.1 Café na nutrição funcional**

O café é uma bebida muito apreciada, especialmente por suas características de sabor e aroma. Além disso, algumas de suas propriedades relacionadas ao bem-estar e à manutenção do estado de alerta também são mencionados pelos autores (MACHADO; DÓREA; COSTA, 2011). No entanto,

café desperta interesse devido aos seus aspectos funcionais associados à saúde. Ainda relatam que é um produto de complexa composição química, principalmente após sua torrefação, durante a qual compostos originais do grão são degradados e novos compostos são formados, garantindo a presença de substâncias nutritivas, bioativas e características de sabor e aroma.

Gomes (2004) relata que a ingestão de café na dieta pode promover redução no índice de peroxidação lipídica e ampliação no teor de glutatona reduzida em ratos normonutridos, assim como em desnutridos. Em pesquisa realizada na Universidade do Arizona com ingestão de café por pessoas idosas, Ryan, Hatfield e Hofstter (2002) mostraram forte ligação entre consumo de café e evolução na performance da memória ao longo do dia, sobretudo verbal e de raciocínio espacial. Já estudos realizados por Smith et al. (2003) indicam que a cafeína interfere no desempenho danoso produzido pela fadiga e aumenta o estado de alerta, com o aumento dos movimentos oculares sacádicos e a ansiedade, desde que ingeridos em pequenas doses.

## **2.2 Análise de medidas repetidas no tempo**

A expressão “medidas repetidas” é utilizada para especificar múltiplas observações da mesma característica em uma ou mais variáveis respostas na mesma unidade experimental (NOBRE; SINGER, 2007).

Os experimentos com medidas repetidas no tempo envolvem dois fatores: tratamentos e tempos, e são frequentes em diversas áreas do conhecimento. O objetivo principal desse tipo de experimento é examinar e comparar as tendências dos tratamentos ao longo do tempo. Isso pode envolver comparações entre tratamentos dentro de cada tempo, ou comparações de tempos dentro de cada tratamento. Dessa forma, o tratamento é o fator entre indivíduos (unidades experimentais) e o tempo o fator intraindivíduos.

Estudos envolvendo medidas repetidas abrangem, entre outros, os delineamentos com parcelas divididas ("split-plot") e delineamentos com intercâmbio ("crossover"), além dos estudos longitudinais (NOBRE; SINGER, 2007). Os esquemas em parcelas divididas são comuns em ciências agrárias, em que há dois estágios na casualização: primeiramente, casualizam-se os níveis de um fator (tratamento) e, em um segundo estágio, há casualização dos níveis de um segundo fator em todos os níveis do primeiro. Já nos planejamentos do tipo "crossover", as unidades experimentais recebem sequências de tratamentos, ou seja, todos os tratamentos são aplicados às mesmas unidades experimentais, após agrupá-las em grupos de número igual ao número de tratamentos.

Faraway (2006) descreve que um planejamento é dito longitudinal quando houver repetição da medida ao longo do tempo, tornando-se, assim, um caso particular de medidas repetidas. O mesmo autor ainda menciona que o principal objetivo no estudo de dados longitudinais é descrever as alterações de uma ou mais variáveis resposta na evolução do tempo e, também, estudar a influência de outros fatores ou covariáveis sobre a variável resposta na unidade experimental, surgindo, assim, centros de interesse da dependência dessa variável ao longo do tempo. Por esta razão, espera-se uma dependência entre as observações referentes a uma mesma unidade experimental.

Em planejamentos longitudinais, as variáveis resposta podem ser contínuas ou discretas, como, por exemplo, ganho de peso e número de folhas. As unidades experimentais podem ser constituídas de um ou mais fatores ou tratamentos e cada uma dessas unidades pode gerar unidades observacionais, em que cada um desses conjuntos de observações pode ser entendido como um perfil individual de resposta para a variável em estudo. O estudo desses perfis individuais pode ser um indício inicial para a inclusão de efeito aleatório no modelo (ROCHA, 2010).

Os dados longitudinais são tidos como regulares quando o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do tempo. E, se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais, tem-se uma estrutura balanceada em relação ao tempo (AMADOR, 2010).

Dados do tipo longitudinais apresentam, conforme Singer, Rocha e Nobre (2004), algumas vantagens, tais como:

- i) permitem estudar o comportamento da variável resposta média da unidade experimental sobre cada tratamento;
- ii) exige menos unidades experimentais (as medidas são feitas nas mesmas parcelas);
- iii) permite verificar existência de dependência da variável resposta em relação às covariáveis existentes;
- iv) diminuição do erro experimental (aumenta do número de dados);
- v) melhora na precisão das estimativas dos parâmetros.

Pela própria obtenção sistemática dos dados longitudinais, é de se esperar que as observações sobre uma mesma unidade experimental tendam a ser correlacionadas. E tal correlação pode ser modelada, explicada por meio da utilização de uma estrutura de covariâncias para os dados observados. E é imprescindível que esse procedimento seja realizado, para que sejam válidas as inferências realizadas (ROCHA, 2010).

Para dados completos, ausência de parcelas perdidas, e balanceados as metodologias de análise de perfil e a análise de curvas de crescimento são técnicas clássicas para análises de dados longitudinais. A análise de perfil pode ser realizada utilizando-se técnicas univariadas ou multivariadas e o ajuste de curvas de crescimento pode ser realizado, além das técnicas anteriores, por meio do modelo linear misto unifatorial que é o foco deste trabalho, no que se refere ao ajuste de curvas para o experimento real aqui tratado.

No estudo de medidas repetidas no tempo, analisadas sob o enfoque da técnica univariada de perfil, ou no esquema de parcela subdividida no tempo, Xavier e Dias (2001) alertam para o fato de os níveis do fator tempo não serem casualizados. E, neste caso, a validade dos resultados da análise de variância é questionável porque acarretará na violação das pressuposições de independências dos erros, pela ausência de casualização e homogeneidade dos erros, pela possível existência de dependência entre as observações mensuradas em uma mesma unidade experimental.

Huynh e Feldt (1970) mostraram que, em um delineamento em esquema de parcelas subdivididas, ainda que os dados sejam longitudinais, com suposição mais geral de que matriz de covariância dos erros possua a forma de simetria composta, o teste F aplicado na análise de variância em relação à subparcela terá distribuição F exata. O teste F aplicado à parcela terá distribuição F exata, independente da forma da matriz de covariância dos erros. Estes autores descreveram uma suposição, ainda mais geral, para a matriz de covariância dos erros, conhecida como condição de Huynh e Feldt (1970), que também mantém distribuição F exata ao teste F aplicado na subparcela.

Entretanto, caso a matriz de covariância não atenda à condição de Huynh e Feldt (1970), mas possua uma forma muito próxima, que será aferida pelo teste de Mauchly, ainda assim as condições do teste F, descritas anteriormente, não se alteram.

Huynh e Feldt (1970) ainda propuseram correções dos graus de liberdade do teste F aplicado na subparcela, necessariamente, se a condição de simetria composta e Huynh e Feldt não forem atendidas. As correções dos graus de liberdade implicarão que o teste F tenha distribuição F aproximada, com isso validando as inferências realizadas na análise de variância.

Outra alternativa para analisar dados mensurados no tempo é a utilização da análise de variância multivariada, também chamada de análise multivariada

de perfil, porque não pressupõe uma estrutura específica para a matriz de covariância dos erros, apenas que os erros tenham distribuição normal multivariada.

Sob o enfoque de ajustar curvas de crescimento, que é muitas vezes de extremo interesse em dados longitudinais, outra forma é pelo uso de modelos lineares mistos, que é uma técnica mais refinada, pois possibilita a utilização de distintas estruturas para as matrizes de variâncias e covariâncias, selecionando-se aquela que melhor se adapte à correlação presente nas unidades experimentais (FREITAS, 2007).

### 2.3 Análise de perfil

Análise de perfil é um tipo de análise de dados longitudinais que pode ser realizada utilizando-se técnicas univariadas e multivariadas e tem como principal objetivo testar hipóteses sobre os valores médios da variável resposta nas diversas ocasiões de observação e compará-los entre os diferentes grupos de unidades experimentais, ou tratamentos (FREITAS et al., 2008).

#### 2.3.1 Análise univariada de perfil

Os modelos para análise univariada de perfis são equivalentes aos costumeiramente considerados em análise de variância (ANOVA), realizada por meio dos planejamentos do tipo parcela subdividida no tempo. Nesses casos, o tempo é tomado como uma variável discreta, ordenando a variável resposta na parcela experimental. O modelo univariado, no esquema de parcelas subdivididas no tempo com o delineamento inteiramente casualizado, é

$$y_{ijk} = \mu + \tau_i + \delta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + e_{ijk} \quad (2.1)$$

em que

$y_{ijk}$  é o valor observado para a variável resposta na  $j$ -ésima unidade experimental submetida ao  $i$ -ésimo tratamento no  $k$ -ésimo tempo;

$\mu$  é uma constante comum a todas as observações;

$\tau_i$  é o efeito do  $i$ -ésimo tratamento;

$\delta_{j(i)}$  representa o efeito aleatório da  $j$ -ésima unidade experimental dentro do  $i$ -ésimo tratamento, tais que  $\delta_{j(i)} \overset{iid}{\cap} N(0, \sigma_\delta^2)$ ;

$\gamma_k$  é o efeito do  $k$ -ésimo tempo observado;

$(\tau\gamma)_{ik}$  é o efeito da interação entre o  $i$ -ésimo tratamento com o  $k$ -ésimo tempo;

$e_{ijk}$  é o erro aleatório associado às observações do  $k$ -ésimo tempo para o  $i$ -ésimo tratamento e  $j$ -ésima parcela experimental, supostos homocedásticos, independentes e normalmente distribuídos, para  $i = 1, \dots, g$  tratamentos,  $j = 1, \dots, b$  unidade experimental e  $k = 1, \dots, t$  tempos, tais que  $e_{ijk} \overset{iid}{\cap} N(0, \sigma_e^2)$ .

Na Tabela 1 é apresentado o esquema da análise de variância, bem como as esperanças dos quadrados médios e o teste F de interesse para o caso balanceado de um delineamento inteiramente casualizado em esquema de parcela subdividida no tempo. Maiores detalhes podem ser obtidos em Storck et al. (2006).

Tabela 1 Esquema para análise da variância, esperanças dos quadrados médios e teste F para as hipóteses de interesse do modelo de experimento no delineamento inteiramente casualizado em esquema de parcelas subdividas no tempo.

Fontes de variação	Graus de liberdade	Quadrados médios	E(QM)	Teste F(sob $H_0$ )
<i>Trat.</i>	$g - 1$	$QM_1$	$\sigma_e^2 + t\sigma_\delta^2 + bt\varphi_g$	$QM_1/QM_2$
<i>Erro (a)</i>	$g(b - 1)$	$QM_2$	$\sigma_e^2 + t\sigma_\delta^2$	
<i>Tempo</i>	$t - 1$	$QM_3$	$\sigma_e^2 + gb\varphi_t$	$QM_3/QM_5$
<i>Trat.×tem</i>	$(g - 1)(t - 1)$	$QM_4$	$\sigma_e^2 + b\varphi_{gt}$	$QM_4/QM_5$
<i>Erro(b)</i>	$(gb - g)(t - 1)$	$QM_5$	$\sigma_e^2$	
<i>Total</i>	$gbt - 1$			

$$\varphi_g = \frac{1}{g-1} \sum_i \tau_i^2, \varphi_t = \frac{1}{t-1} \sum_k \gamma_k^2 \text{ e } \varphi_{gt} = \frac{1}{(t-1)(g-1)} \sum_{ik} (\tau\gamma)_{ik}^2.$$

As simbologias *Trat.*, *tem*, e *QM*, na Tabela, implicam em tratamento, tempo e quadrado médio, respectivamente. Os quadrados médios são obtidos dividindo-se as respectivas somas de quadrados pelos graus de liberdade, cujas fórmulas encontram-se a seguir:

$$SQ_1 = QM_1(g - 1), \quad SQ_2 = QM_2(g(b - 1)), \quad SQ_3 = QM_3(t - 1), \\ SQ_4 = QM_4(g - 1)(t - 1) \text{ e } SQ_5 = QM_5(gb - g)(t - 1).$$

E também

$$C = \frac{y_{..}^2}{gbt}, \quad SQ_1 = \frac{1}{bt} \sum_{i=1}^g y_{i..}^2 - C, \quad SQ_2 = \frac{1}{t} \sum_{i=1}^g \sum_{j=1}^b y_{ij.}^2 - \frac{1}{bt} \sum_{i=1}^g y_{i..}^2,$$

$$SQ_3 = \frac{1}{gb} \sum_{k=1}^t y_{.k}^2 - C, \quad SQ_4 = \frac{1}{b} \sum_{i=1}^g \sum_{k=1}^t y_{i.k}^2 - C - SQ_1 - SQ_3,$$

$$SQ_5 = SQ_6 - SQ_4 - SQ_3 - SQ_2 - SQ_1 \quad e \quad SQ_6 = \sum_{i=1}^g \sum_{j=1}^b \sum_{k=1}^t y_{ijk}^2 - C.$$

As hipóteses de interesse a serem testadas são as seguintes:

- 1) para efeito de tratamentos, conforme Storck et al. (2006),

$$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$$

$$H_1: \text{pelo menos um } \tau_i \neq 0;$$

- 2) para efeito de tempo:

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_t = 0$$

$$H_1: \text{pelo menos um } \gamma_k \neq 0;$$

- 2) para efeito da interação tratamento x tempo:

$$H_0: (\tau\gamma)_{11} = (\tau\gamma)_{12} = \dots = (\tau\gamma)_{gt} = 0$$

$$H_1: \text{pelo menos um } (\tau\gamma)_{ik} \neq 0.$$

É comum considerar o nível mínimo de significância igual a 0,05 para testar as hipóteses de nulidades. Sempre que o valor da probabilidade do teste F for menor ou igual a 0,05, rejeita-se  $H_0$  e há implicação de que pelo menos os efeitos de dois níveis são diferentes, em média. A averiguação dos resultados é iniciada pelas interações duplas, procedendo-se aos desdobramentos em caso de rejeição da hipótese nula. E, em caso da aceitação da hipótese nula para este caso, o passo seguinte será o estudo dos efeitos principais.

Cecon et al. (2008) descreveram uma diferenciação entre a análise de medidas repetidas e a parcela subdividida no tempo. Na segunda, assume-se que as respostas de diferentes tempos na mesma parcela experimental são igualmente correlacionadas. Já na primeira, as respostas em tempos mais próximos, também na mesma parcela experimental, podem ser mais fortemente correlacionadas, comparadas à de tempos mais distantes.

Nos casos de esquemas de medidas repetidas, não apenas no tempo, unifatoriais, Montgomery (2001) propõe uma análise de variância com

particularidades distintas da usual. Essa diferenciação é para que, possivelmente, o erro experimental não seja inflacionado pela correlação existente entre as observações de uma mesma unidade experimental. A análise é construída obtendo-se o erro experimental, sem considerar a soma dos quadrados intraunidade experimental.

De acordo com Xavier e Dias (2001), o esquema em parcela subdividida no tempo pressupõe que a estrutura da matriz de covariância seja do tipo uniforme, no qual as variâncias são iguais nos diversos tempos e covariâncias iguais entre dois instantes de tempo quaisquer. Já sob o enfoque dado por Littell et al. (2006), em experimentos utilizando medidas repetidas no tempo, é possível que a estrutura da matriz de covariância se diferencie, impondo uma estrutura de correlação distinta da descrita inicialmente.

Loughin et al. (2007) fizeram consideráveis críticas aos modelos univariados que utilizam medidas repetidas ao longo do tempo, por não considerar as correlações existentes nas parcelas experimentais, e isso pode invalidar todos os resultados das análises estatísticas. Ainda relatam que, se não for considerada a dependência entre as observações na unidade experimental, estudos de simulação demonstraram que haverá elevada taxa do erro tipo I e que todos os componentes de variância estão distorcidos da realidade. Sendo assim, o estudo proposto pelos autores converge para técnicas de modelagem que possibilite efetivar ampla discussão sobre os efeitos dos parâmetros inseridos no modelo e que o mais plausível é aquele em que a dependência das observações na mesma unidade experimental possa ser modelada por meio da inclusão de fatores de efeitos aleatórios no modelo.

Vieira et al. (2007) citam que uma condição suficiente para que o teste F da análise de variância usual, quando se utilizam esquemas de parcelas subdivididas, seja exato, em nível de subparcela necessariamente, é que a matriz

de covariâncias dos erros ( $\Sigma$ ) tenha uma forma chamada de simetria composta, ou seja, uma matriz uniforme do tipo

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 \end{bmatrix} \quad (2.2)$$

em que  $\sigma^2 = \sigma_\pi^2 + \sigma_e^2$ , como descrito anteriormente,  $\sigma_\pi^2$  é a variância da parcela e  $\sigma_e^2$  a variância da subparcela e  $\sigma_1 = \sigma_\pi^2$ .

Huynh e Feldt (1970) descreveram uma condição necessária e suficiente mais geral da forma de  $\Sigma$ , que também garante que os testes F referentes à subparcela sejam válidos. Tal condição, simbolizada por (H-F) de Huynh-Feldt, especifica que os elementos da matriz de covariâncias sejam expressos, para um  $\lambda > 0$ , da seguinte maneira:

$$\begin{aligned} \sigma_{ii} &= \sigma_i^2 \\ \sigma_{ij} &= [(\sigma_i^2 - \sigma_j^2)/2] - \lambda \end{aligned} \quad (2.3)$$

ou alternativamente,

$$\text{var}(x_i - x_j) = 2\lambda; \text{ para } i \neq j \quad (2.4)$$

em que  $\lambda$  é a diferença entre a média das variâncias e a média das covariâncias. Sendo assim, essa condição é equivalente à de simetria composta, quando todas as variâncias forem iguais. Nota-se que matrizes de covariâncias que possuam variâncias iguais e correlações nulas são casos especiais da condição (H-F) e também vê-se que é uma condição menos restritiva que a condição de simetria composta. A matriz que atenda à condição (H-F) é denominada matriz circular. Cabe ressaltar que toda matriz que possua simetria composta é circular, mas a recíproca não é verdadeira (ENDE, 1993).

Pode ocorrer que a matriz de covariância dos erros não atenda à condição numérica imposta pela condição (H-F), porém, pode ser que haja uma aproximação estatística. Sendo assim, Huynh e Feldt (1970) propuseram um teste de esfericidade a partir do teste de Mauchly (1940) para verificar se a matriz de covariâncias  $\Sigma^*$  satisfaz à condição de esfericidade. Os mesmos autores provaram que, se  $\Sigma$  atender à condição H-F, então,  $\Sigma^*$  pode ser escrita como

$$\Sigma^* = C\Sigma C' = \lambda I_{t-1} \quad (2.5)$$

em que  $C_{(t-1) \times t}$  é uma matriz de dimensões  $((t-1) \times t)$ , que pode ser construída como alguma matriz que define os coeficientes de  $t-1$  contrastes ortonormais entre as  $t$  medidas de tempo; conseqüentemente,  $\lambda$  é uma constante maior que 0 e  $I$  é a matriz identidade, de dimensões  $(t-1) \times (t-1)$ .

Para verificar a igualdade, Huynh e Feldt (1970) descreveram um teste a partir do teste de esfericidade de Mauchly.

### 2.3.1.1 Teste de esfericidade de Mauchly

O teste de esfericidade de Mauchly consiste em verificar se uma população normal multivariada possui variâncias iguais e correlações nulas. Se essa condição for atendida, a matriz de covariâncias será denominada de esférica (MAUCHLY, 1940).

Para a sua aplicação, como foi descrito por Huynh e Feldt (1970), é utilizada uma matriz de contrastes ortonormais  $C$  que consiste em transformar a matriz de covariância  $\Sigma$  original para uma forma ortonormalizada  $\Sigma^*$ , cuja esfericidade será aferida pelo teste e, caso  $\Sigma^*$  seja tida como esférica, a matriz original  $\Sigma$ , atenderá à condição H-F e implicará na garantia que os testes  $F$  da subparcela, no caso da parcela subdividida no tempo possuirá distribuição  $F$  exata. A seguir, a descrição do teste.

Sob a hipótese nula,

$$H_0: C\Sigma C' = \lambda I_{t-1}$$

que é a hipótese de que a matriz de covariância satisfaz à condição de esfericidade, a estatística  $W$  do teste Mauchly (1940) poderá ser utilizada para testar esta hipótese de independência e homocedasticidade de uma variável aleatória normal multivariada. Se  $S$  é a matriz de covariância amostral para o erro experimental (intraindivíduo) da subparcela e  $p = t - 1$ , o número de contrastes ortogonais, então, o valor de  $W$  é dado

$$W = \frac{|CSC'|}{[tr CSC'/p]^p} \quad (2.6)$$

sendo que o valor de  $W$  não depende da escolha particular para esta matriz de transformação  $C$ . Se  $\Sigma$  é estimada com  $n_1$  graus de liberdade e se considerar,

$$f = [p(p+1)/2] - 1$$

$$d = 1 - (2p^2 + p + 2)/6pn_1$$

então,  $n_1 d \ln W$  é, aproximadamente, distribuído por uma distribuição qui-quadrado central com  $f$  graus de liberdade, ou seja,  $-n_1 d \ln W > \chi_{\alpha, f}^2$ , rejeita-se a hipótese nula a  $\alpha$  de significância.

Todavia, Freitas et al. (2008) não recomendam o uso deste teste quando há dúvidas da pressuposição de normalidade dos erros, devido à sua sensibilidade a ausência de normalidade e, com isso, acarreta a diminuição de seu poder. Sendo assim, é aconselhável a utilização de testes de normalidade para aferir essa pressuposição, assim como uma ampla análise exploratória.

Em alguns trabalhos, especialmente em ciências agrárias, não utilizam a técnica univariada de perfil caso o teste não indique a esfericidade da matriz de covariância. Aparício et al. (2010), por exemplo, avaliaram, em experimento, o crescimento inicial de dois clones do eucalipto e verificaram que o teste rejeitou a hipótese nula de esfericidade da matriz de covariância e, por esse fato, optaram pela não realização da análise univariada de perfil.

Malheiros (2004), estudando, via simulação, a esfericidade da matriz de covariância dos erros em experimentos com medidas repetidas no tempo, demonstrou que os valores da estatística que aferem a significância da esfericidade para estruturas que satisfazem a este critério foram próximos, entretanto, superiores aos valores obtidos para estruturas que não atendem ao critério de esfericidade, isso utilizando dados balanceados. Já utilizando dados desbalanceados, houve menos precisão comparado a dados balanceados, porém, ainda assim, os valores da estatística que utilizaram estruturas que satisfazem ao critério de esfericidade foi superior as demais.

Quando utilizar a análise univariada em esquema de parcelas subdivididas no tempo e a matriz de covariância dos erros intraindivíduos não possuir, ao menos, a forma de H-F, uma alternativa é proceder ao ajuste dos graus de liberdade do teste F para os fatores presentes na subparcela

### 2.3.1.2 Correções do número de graus de liberdade

Autores como Box (1954), Geisser e Greenhouse (1958) e Huynh e Feldt (1976) sugerem algumas correções nos graus de liberdade para o fator tempo e para a interação entre tratamentos e tempos, possibilitando que esses testes possam ser usados, ainda que de forma aproximada, mesmo que a condição de esfericidade da matriz  $\Sigma$  não seja satisfeita. As correções propostas pelos respectivos autores são:

1) ajuste de Geisser e Greenhouse (1958), dado por

$$\hat{\varepsilon} = \frac{[\text{tr}(\text{CSC}')]^2}{(t-1)\text{tr}(\text{CSC}')^2} \quad (2.7)$$

ou, ainda, admitindo  $A_{qq} = CSC'$ , com  $q = t - 1$  contrastes ortogonais normalizados, sobre  $t$  medidas repetidas e  $a_{ii}$ , com  $i = 1, 2, \dots, q$ , as raízes latentes de  $A_{qq}$ , o fator de correção pode ser alterado por

$$\hat{\varepsilon} = \frac{(\sum_{i=1}^q a_{ii})^2}{(t-1) \sum_{i=1}^q \sum_{j=1}^q a_{ij}^2} \quad (2.8);$$

2) ajuste de Huynh e Feldt (1976), dado por

$$\tilde{\varepsilon} = \frac{[N(t-1)\hat{\varepsilon} - 2]}{(t-1)[N - g - (t-1)\hat{\varepsilon}]} \quad (2.9)$$

em que  $N$  é o número total de indivíduos,  $g$  é o número de níveis do fator da parcela e  $t$  é o número de medidas repetidas. Box (1954) demonstrou que o ajuste extremo que pode ser utilizado quando a matriz de covariância não possuir o formato esférico é  $\varepsilon = (t-1)^{-1}$ . Ainda demonstra que  $(t-1)^{-1} \leq \varepsilon \leq 1$ , isto é, o valor máximo de 1 corresponde à condição de esfericidade e, à medida que a matriz de covariância  $\Sigma$  se afasta do padrão, o valor de  $\varepsilon$  tende ao seu limite inferior. É demonstrável que, se a condição de esfericidade for atingida,  $\varepsilon = 1$ . Nota-se também que  $\tilde{\varepsilon}$  pode assumir valores maiores que 1 e, neste caso, deve adotá-lo igual a 1. Entretanto, quando as estimativas de  $\varepsilon$  estiverem próximas dos limites, isto induzirá a testes muito conservadores e liberais, respectivamente.

Malheiros (2001) verificou, via estudo de simulação, que os testes F da análise da variância são imprecisos com a correção de Huynh e Feldt (1976) e com a correção de Greenhouse e Geisser (1958) são precisos, independente dos dados serem balanceados e da estrutura da matriz  $\Sigma$ . Huynh e Feldt (1976) afirmam que  $\hat{\varepsilon}$  tem a desvantagem de superestimar o verdadeiro valor do nível de significância e, ainda, que, em geral,  $\hat{\varepsilon}$  parece ser melhor estimador comparado a  $\tilde{\varepsilon}$ , quando o real valor de  $\varepsilon$  está ao redor de 0,5, ou menor. Sendo  $\hat{\varepsilon} = 1$ , Huynh e Feldt (1976) propuseram que a condição (H-F) para uma matriz de covariância  $\Sigma$  está atendida.

Quando a condição de esfericidade da matriz de covariância não é satisfeita uma alternativa, além da correção dos graus números de liberdade, é a análise multivariada de perfil.

### **2.3.2 Análise multivariada de perfil**

Como já comentado em tópico anterior, a suposição de que as observações são independentes em dados provenientes de medidas repetidas não é adequada, pois as medidas tomadas em um mesmo indivíduo tendem a ser correlacionadas. Portanto, em casos desse tipo, uma solução natural para análise de medidas repetidas no tempo é a análise de variância multivariada dita MANOVA, também conhecida como análise multivariada de perfil (ANDRADE; SINGER, 1986; VIEIRA et al., 2007).

A denominação de análise multivariada de perfil deve-se ao fato de serem analisadas simultaneamente as relações entre as medidas múltiplas temporais da variável tempo para cada indivíduo (unidade experimental), com alguma outra variável, usualmente denominada de tratamento.

Ainda segundo os mesmos autores, diferentemente do modelo de análise de variância (ANOVA), o modelo MANOVA não requer que a variância das medidas repetidas ou que a correlação entre os pares de medidas repetidas permaneçam constantes ao longo do tempo, porém, ambos os modelos requerem que as variâncias e as correlações sejam homogêneas a cada instante do tempo.

Hair et al. (2009) e Meredith e Stehrnan (1991) ressaltam que a análise multivariada não impõe nenhuma restrição ou condição para a matriz de covariância, e que também não há suposição sobre a estrutura da matriz de covariâncias, tornando-se, portanto, uma solução plausível para conjunto de dados longitudinais. Sendo assim, o enfoque multivariado possibilita a opção de um modelo bastante geral, no que se refere a representar a estrutura de

covariâncias, admitindo-se que as variâncias e as covariâncias entre tempos distintos sejam diferentes.

Alguns autores, como Hair et al. (2009), afirmam que o modelo multivariado tem uma vantagem muito importante de permitir a adoção de um modelo bastante geral para representar a estrutura de covariância, como citado anteriormente. Entretanto, apresenta a desvantagem de necessitar de perfis individuais completos e também o baixo poder dos testes, devidos ao excesso de parametrização. Porém, com exceção das restrições citadas anteriormente, em boa parte dos casos, a análise de variância multivariada será uma técnica mais adequada.

O modelo utilizado por Freitas et al. (2007), em experimento com cana-de-açúcar em esquema de parcelas subdividida no tempo, adequando-o para um delineamento inteiramente casualizado com medidas repetidas no tempo, foi

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk} \quad (2.10)$$

em que

$y_{ijk}$  é o valor observado da variável resposta no  $i$ -ésimo tratamento,  $j$ -ésimo tempo na  $k$ -ésima repetição;

$\mu$  é uma constante comum a todas as observações;

$\tau_i$  é o efeito do  $i$ -ésimo tratamento;

$\gamma_j$  é o efeito do  $j$ -ésimo tempo;

$(\tau\gamma)_{ij}$  é o efeito da interação entre o  $i$ -ésimo tratamento com  $j$ -ésimo tempo;

$e_{ijk}$  é o erro aleatório associado à observação do  $i$ -ésimo tratamento,  $j$ -ésimo tempo na  $k$ -ésima repetição, supostos homocedásticos, independentes e normalmente distribuídos, para  $i = 1, \dots, g$  tratamentos,  $j = 1, \dots, t$  tempos e  $k = 1, 2, \dots, K$  repetição.

No modelo (2.10) verifica-se que não foi considerado ou incluído o erro da parcela. Os modelos de delineamento experimentais, em geral, apresentam várias soluções para os parâmetros, constituindo um sistema indeterminado

(MONTGOMERY, 2001). Uma das alternativas é a utilização de restrições, com a finalidade de se ter uma matriz de posto completo e o sistema ser determinado (SEARLE, 1971).

Um conjunto de restrições paramétricas, de acordo com Montgomery (2001), é:

$$\sum_{i=1}^g \tau_i = \sum_{j=1}^t \gamma_j = \sum_{i=1}^g (\tau\gamma)_{ij} = \sum_{j=1}^t (\tau\gamma)_{ij} = 0 \quad (2.11)$$

Com o enfoque multivariado, o vetor de erro aleatório  $e_{ik} = [e_{ik1}, \dots, e_{ikt}]'$  tem  $e_{ik} \sim N_t(0, \Sigma)$ , em que  $\Sigma$  é uma matriz  $t \times t$ , com estrutura geral.

O modelo utilizado na análise multivariada de perfil (MANOVA) não se distingue dos modelos usuais utilizados por autores como Ferreira (2008) e Xavier e Dias (2001), ao estudarem a análise de variância multivariada. Em sua forma matricial, sua representação é

$$Y = XB + E \quad (2.12)$$

sendo  $Y$  a matriz dos dados observados de  $t$  variáveis respostas para as  $n$  unidades experimentais de dimensões  $(n \times t)$  representada por

$$Y = \begin{bmatrix} y_{111} & y_{112} & \dots & y_{11t} \\ y_{121} & y_{122} & \dots & y_{12t} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1k1} & y_{1k2} & \dots & y_{1kt} \\ \vdots & \vdots & \ddots & \vdots \\ y_{g11} & y_{g12} & \dots & y_{g1t} \\ \vdots & \vdots & \ddots & \vdots \\ y_{gk1} & y_{gk2} & \dots & y_{gkt} \end{bmatrix}; \quad (2.13)$$

$X$  é a matriz de delineamento conhecida, de dimensões  $(gk \times (g + 1))$ , para o delineamento inteiramente casualizado com  $g$  tratamentos e  $k$  repetições, a qual pode ser decomposta em:

$$X = [X_1 \ : \ X_2], \text{ com } X_1 = [1_{gk}], X_2 = \begin{bmatrix} 1_k & 0 & \dots & 0 \\ 0 & 1_k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1_k \end{bmatrix}$$

com  $1_{gk}$  e  $1_k$  vetores de elementos iguais a um com dimensões  $(gk \times 1)$  e  $(k \times 1)$ , respectivamente;

$B$  é a matriz dos parâmetros, de dimensões  $((g + 1) \times t)$ , dada por

$$B = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_t \\ \tau_{11} & \tau_{12} & \dots & \tau_{1t} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{g1} & \tau_{g2} & \dots & \tau_{gt} \end{bmatrix} \text{ e}$$

$E$  é a matriz dos erros experimentais, de dimensões  $(n \times t)$ , e representada por

$$E = \begin{bmatrix} \epsilon_{111} & \epsilon_{112} & \dots & \epsilon_{11t} \\ \epsilon_{121} & \epsilon_{122} & \dots & \epsilon_{12t} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{1k1} & \epsilon_{1k2} & \dots & \epsilon_{1kt} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{g11} & \epsilon_{g12} & \dots & \epsilon_{g1t} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{gk1} & \epsilon_{gk2} & \dots & \epsilon_{gkt} \end{bmatrix} = \begin{bmatrix} \epsilon'_{11} \\ \epsilon'_{12} \\ \vdots \\ \epsilon'_{1k} \\ \vdots \\ \epsilon'_{g1} \\ \vdots \\ \epsilon'_{gk} \end{bmatrix} \quad (2.14)$$

O vetor de erro experimental  $\epsilon_{ij} = [\epsilon_{ij1} \ \epsilon_{ij2} \ \dots \ \epsilon_{ijt}]'$  está associado à  $ij$ -ésima unidade experimental. Sendo assim, o vetor  $\epsilon_{ij}$  associado

às linhas da matriz  $E$  e correspondente à observação  $y_{ij}$  é considerado como tendo distribuição normal  $t$ -variada, com vetor de média  $\Phi$  e matriz de covariância  $\Sigma$ , representada por:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1t} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{t1} & \sigma_{t2} & \dots & \sigma_t^2 \end{bmatrix} \quad (2.15)$$

Por extensão, os perfis de respostas  $y_{ij}$  apresentam distribuições normais  $t$ -variadas com média

$$E[y_{ij}] = XB$$

e variância

$$V[y_{ij}] = \Sigma.$$

Logo,  $y_{ij} \sim N_t(XB, \Sigma)$  e a matriz de covariância  $\Sigma$  é não estruturada e tem  $t(t + 1)/2$  parâmetros.

Para situações como esta, podem-se avaliar três hipóteses de interesse (FREITAS, 2007), representadas por:

i) Os perfis médios de respostas associados aos tratamentos são paralelos, indicando que não há interação entre os fatores que definem os tratamentos e os tempos mensurados. Esta hipótese deve ser a primeira a ser testada, pois sua aceitação determinará a forma das outras hipóteses de interesse.

A representação na forma matricial é

$$H_{0I}: \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(t-1)} - \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(t-1)} - \mu_{2t} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} - \mu_{g2} \\ \mu_{g2} - \mu_{g3} \\ \vdots \\ \mu_{g(t-1)} - \mu_{gt} \end{bmatrix}$$

ii) Os perfis médios de respostas referentes aos tratamentos são coincidentes, indicando que não há efeito desse fator. Em relação aos parâmetros dos modelos, a hipótese pode ser representada por:

$$H_{0G}: \sum_{k=1}^t \mu_{1k} = \sum_{k=1}^t \mu_{2k} = \dots = \sum_{k=1}^t \mu_{gk}$$

ou em sua forma matricial, tem-se:

$$H_{0G}: \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2t} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} \\ \mu_{g2} \\ \vdots \\ \mu_{gt} \end{bmatrix}$$

iii) Os perfis médios de respostas referentes aos tratamentos são paralelos ao eixo das abscissas, indicando que não há efeito do fator tempo. Em relação aos parâmetros dos modelos, a hipótese pode ser representada por:

$$H_{0T}: \sum_{i=1}^g \mu_{i1} = \sum_{i=1}^g \mu_{i2} = \dots = \sum_{i=1}^g \mu_{it}$$

Ou, em sua forma matricial, tem-se:

$$H_{0T}: \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{g2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{gt} \end{bmatrix}.$$

No intuito de unificação de resultados, qualquer hipótese pode ser expressa a partir de uma forma de hipótese linear geral (ANDRADE; SINGER, 1986; FREITAS, 2007):

$$H_0 = CBM = 0$$

em que  $C_{((g-1) \times (g+1))}$  e  $M_{((t) \times (t-1))}$  são matrizes com posto  $g$  e  $t$ , respectivamente e  $B$ , a matriz os parâmetros citada no modelo (2.12). A matriz  $C$  é utilizada para obter comparações entre funções estimáveis no âmbito dos tratamentos, que são as linhas da matriz  $B$ . Já a matriz  $M$  é responsável por

comparações entre os tempos. As formas para as matrizes C e M, no que tange às hipóteses de interesse, são:

$$H_{0I}: C_1 ((g-1) \times (g+1)) = \begin{bmatrix} 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

e

$$M_1 (t \times (t-1)) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{bmatrix};$$

$H_{0G}: C_1$  e  $M_2 = 1_t$  e  $H_{0T}: C_2 = 1'_g$  e  $M_1$  sendo  $1_t$  e  $1'_g$  vetores de uns com dimensões  $t$  e  $g$ , respectivamente.

Para que seja testada a hipótese linear geral, segundo Ferreira (2008), os testes consideram as matrizes somas de quadrados e produtos de tratamento e do erro, bem como a utilização do conceito de variância generalizada. Esses testes podem ser obtidos por distintos critérios. Segundo Freitas (2007), em geral, suas estatísticas de teste são obtidas a partir das raízes latentes da matriz  $HE^{-1}$ , sendo que

$$H = M'[(X'X)^{-1}X'Y]C'[C(X'X)^{-1}C']^{-1}C[(X'X)^{-1}X'Y]M$$

é a matriz de somas de quadrados e produtos relativas a hipótese nula e

$$E = M'Y'[I - X(X'X)^{-1}X']^{-1}YM$$

é a matriz de somas de quadrados e produtos relativas ao erro, conforme Vieira et al. (2007).

Para avaliação das hipóteses, na literatura há várias opções. Segundo Ferreira (2008) e Johnson e Wichern (2007), entre outros, estatísticas de testes multivariados são obtidas por meio dos princípios da união-intersecção de Roy e da razão de verossimilhança de Wilks. As principais estatísticas são:

1) Lambda de Wilks

$$\Lambda = \prod_{i=1}^s \frac{1}{1 + \lambda_i} \quad (2.16)$$

2) Raiz máxima de Roy

$$\theta_s = \max(\lambda_i / (1 + \lambda_i)) \quad (2.17)$$

3) Traço de Hotelling-Lawley

$$T = \sum_{i=1}^s \lambda_i \quad (2.18)$$

4) Traço de Pillai

$$P = \sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i}. \quad (2.19)$$

Nas quais  $\lambda_i$  é a  $i$ -ésima raiz latente de  $HE^{-1}$  e  $S$  é o número das raízes latentes não nulas, sendo  $S = \min(t - 1; g - 1)$ . As distribuições exatas dessas estatísticas dependem, sob hipótese nula, dos parâmetros  $m_1 = (|(t - 1) - (g - 1)| - 1)/2$  e  $m_2 = (n - g - (t - 1) - 1)/2$ .

Segundo Johnson e Wichern (2007), os quatro testes multivariados podem produzir diferentes níveis descritivos. A ordem de preferência, em termos de poder, é: Traço de Pillai, Lambda de Wilks, Traço de Lawley-Hotelling e a Maior Raiz de Roy (VIEIRA et al., 2007). Entretanto, Ferreira (2008) salienta que o Critério de Wilks é o mais utilizado, por razões históricas, embora seu poder não seja superior ao dos demais. Johnson e Wichern (2007) salientam que

o Traço de Pillai é mais robusto a não normalidade dos dados, comparado aos demais.

Quando a condição de esfericidade da matriz de covariância não é satisfeita, além da análise multivariada de perfil, uma alternativa é a análise de curvas de crescimento por meio de modelos lineares mistos, estudada por vários autores, como Henderson (1949, 1975) e Potthoff e Roy (1964).

#### **2.4 Modelo linear misto, com referência a curvas de crescimento**

Nos experimentos em que as avaliações são realizadas nas mesmas unidades experimentais ao longo do tempo é de esperar que ocorra uma correlação entre as medidas no tempo e exista certa heterogeneidade de variâncias (LITTELL et al., 2006). Uma consequência de se ignorar a presença da correlação e ou heterogeneidade das variâncias entre as medidas no tempo é que a significância entre as medidas dos tratamentos pode ser exagerada e a sensibilidade dos testes para a interação é reduzida.

Em situações como essa, uma abordagem que pode ser utilizada é a modelagem com modelos mistos. Pinheiro e Bates (2000) definem como modelo misto aquele no qual estão presentes efeitos fixos e aleatórios. Sendo assim, um modelo linear misto é aquele que apresenta fatores ou parâmetros de efeitos fixos e de efeitos aleatórios, além do erro experimental e da constante  $\mu$  (LITTELL et al., 2006). A inclusão de efeitos aleatórios possibilita a modelagem da correlação intraindivíduo, presente, muitas vezes, em dados agrupados e aos quais se permite atribuir uma distribuição de probabilidade. Portanto, o modelo misto torna-se uma extensão do modelo linear geral e apresenta maior flexibilidade porque permite a inclusão de estrutura de interdependência das observações e, com isso, acomoda a estrutura de correlação presente nas medidas repetidas.

Os modelos lineares mistos têm as mais variadas aplicações e tornaram-se mais conhecidos nas últimas décadas, e têm sido tema de grande importância na área de estatística. Essa metodologia foi inicialmente estudada a partir da metade do século XX, por Henderson (1949) e tornou-se mais conhecida nas últimas décadas devido ao advento de técnicas computacionais robustas, pois, devido a esse fato, até então se usavam, predominantemente, modelos de efeitos fixos.

A utilização de modelos lineares mistos no intuito de modelar curvas polinomiais de crescimento torna-se uma técnica alternativa versátil, pois possibilita considerar distintas estruturas para a matriz de variâncias e covariâncias. Segundo Laird e Ware (1982), tem a vantagem de poder ser utilizado em conjunto de dados não balanceados em relação ao tempo e, ainda, permite a inclusão da matriz de covariância com número menor de parâmetros, comparado aos modelos multivariados.

Os modelos lineares mistos fazem a junção de diversas análises estatísticas, tornando-se eficazes para análise de planejamentos longitudinais, parcelas subdivididas e modelos que utilizam variâncias heterogêneas. Nos modelos de curvas de crescimento, o interesse básico está na estimação e na predição dos efeitos de tratamentos em algum tempo, por exemplo. Diferenciando-se do modelo usual de medidas repetidas no tempo, no qual o interesse é diagnosticar diferenças nos efeitos dos tratamentos ao longo do tempo.

McCulloch e Searle (2001) sugerem a inclusão de variáveis latentes em modelos lineares mistos como alternativa para modelar a estrutura de correlação intraunidades experimentais. A variável latente é a variável que não pode ser medida diretamente, como, por exemplo, inteligência, nível sócio-econômico, qualidade de vida e também outras características do modelo, como os efeitos aleatórios citados pelos autores.

No ajuste de curvas de crescimento, quando há interesse em encontrar uma relação entre a variável resposta e o tempo, os coeficientes de regressão das curvas individuais serão considerados efeitos aleatórios, variando de indivíduo para indivíduo e os coeficientes de regressão das curvas médias, efeitos fixos (HAUSER et al., 2009).

O modelo linear misto em dois estágios foi estudado por autores como Laird e Ware (1982) e Littell et al. (2006), dentre outros. Eles consideraram os efeitos fixos no primeiro estágio para a obtenção da curva polinomial média e, no segundo estágio, a inclusão dos efeitos aleatórios para a obtenção das diferentes curvas para cada indivíduo.

A especificação do modelo linear misto pode ser feita na seguinte forma matricial (HENDERSON, 1975):

$$y = X\beta + Zu + e \quad (2.20)$$

em que

$y$  representa o vetor das respostas individuais ou vetor de observações, de dimensões  $n \times 1$ ;

$X$  é uma matriz do planejamento experimental relativa aos parâmetros de efeitos fixos, de dimensões  $n \times r$ , e posto  $r < n$ ;

$\beta$  é o vetor de parâmetros fixos, desconhecidos, de dimensão  $r \times 1$ ;

$Z$  é uma matriz conhecida do planejamento experimental, relativa aos efeitos aleatórios, de dimensões  $n \times q$ ;

$u$  é o vetor de efeitos aleatórios, desconhecidos,  $q \times 1$ ;

$e$  é o vetor de erros aleatórios não observáveis, de dimensão  $n \times 1$ .

As matrizes  $X$  e  $Z$  podem se diferenciar, podendo  $Z$  conter qualquer covariável que influencie a unidade experimental. A formulação de  $X$  é semelhante à utilizada na análise usual de regressão, em que suas colunas especificam os fatores que definem a estrutura das subpopulações (tratamentos)

ao fator tempo, identificando a curva a ser ajustada e as covariáveis cujos efeitos na resposta desejam-se obter.

No modelo (2.20) assume-se que  $E[u] = 0$  e  $E[e] = 0$  com matrizes de covariâncias representadas por

$$V \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix},$$

com suposição que  $u$  e  $e$  são independentes e normalmente distribuídos, então,  $u \sim N(0, G)$  e  $e \sim N(0, R)$ .

O modelo linear misto é usualmente especificado em termos das respostas condicionais aos efeitos aleatórios, sendo  $y$  um vetor específico de medidas repetidas para alguma unidade experimental

$$y|u \sim NM(X\beta + Zu, R).$$

Entretanto, as inferências são realizadas no modelo marginal, portanto,

$$V[y] = V[X\beta + Zu + e]$$

$$V[y] = V[Zu + e], \text{ pois } V[X\beta] = COV[X\beta, e] = COV[e, X\beta] = \Phi$$

$$V[y] = V[Zu] + V[e] + COV[Zu, e] + COV[e, Zu]$$

$$V[y] = ZV[u]Z' + R + ZCOV[u, e] + COV[e, u]Z'$$

Com relação às propriedades do modelo (2.20), mostrou-se que

$$V[y] = ZGZ' + R + Z\Phi_{(q \times n)} + \Phi_{(n \times q)}Z'$$

logo,

$$V[y] = ZGZ' + R$$

e que

$$E[y] = E[X\beta + Zu + e] = E[X\beta] + E[Zu] + E[e] = X\beta$$

assim,  $y \sim NM(X\beta, ZGZ' + R)$  ou  $y \sim NM(X\beta, V)$ .

### 2.4.1 Estimação do parâmetro $\beta$ e predição de $u$

Nesta classe de modelos há o interesse prático de estimação dos efeitos fixos, predição dos efeitos aleatórios e estimação dos componentes de variância. A obtenção das equações de modelos mistos é feita pela minimização da soma de quadrados dos resíduos ou pela maximização da função densidade de probabilidade conjunta de  $y$  e  $u$  (FREITAS, 2007).

A função densidade de probabilidade de  $y$  é escrita como

$$f(y) = \frac{1}{2\pi^{n/2}(|ZGZ' + R|)^{1/2}} e^{-1/2[(y - X\beta)'(ZGZ' + R)^{-1}(y - X\beta)]}$$

podendo a função de probabilidade conjunta de  $y$  e  $u$  ser escrita como

$$f(y, u) = f(y|u) \cdot f(u)$$

Dessa forma, tem-se:

$$f(y, u) = \frac{1}{(2\pi)^{n/2}(|R|)^{1/2}} e^{-1/2[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu)]} \cdot \frac{1}{(2\pi)^{j/2}(|G|)^{1/2}} e^{-1/2[(u - 0)'G^{-1}(u - 0)]}$$

em que  $n$  e  $j$  são os postos das matrizes  $R$  e  $G$ , respectivamente.

No intuito de proceder à maximização da função de densidade de probabilidade conjunta, para efeito de simplificação, usa-se trabalhar com a função suporte  $l$ , ou seja, utiliza-se o logaritmo neperiano da função de verossimilhança, isto é:

$$l(\beta, u|y) = \frac{(n+j)}{2} \log(2\pi) - \frac{1}{2} (\log R + \log G) - \frac{1}{2} (y'R^{-1}y - 2y'R^{-1}X\beta +$$

$$- 2y'R^{-1}Zu + 2\beta'X'R^{-1}Zu + \beta'X'R^{-1}X\beta + u'Z'R^{-1}Zu + u'G^{-1}u).$$

Derivando  $l(\beta, u|y)$  em relação a  $\beta$  e  $u$ , tomando-se as derivadas parciais iguais a 0, obtém-se:

$$\begin{bmatrix} \frac{\partial l(\beta, u | y)}{\partial \beta} \\ \frac{\partial l(\beta, u | y)}{\partial u} \end{bmatrix} = \begin{bmatrix} -X'R^{-1}y + X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{u} \\ -Z'R^{-1}y + Z'R^{-1}X\hat{\beta} + Z'R^{-1}Z\hat{u} + G^{-1}\hat{u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Desenvolvendo-se as expressões e agrupando-se de modo conveniente, tem-se

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}.$$

Essas equações são conhecidas na literatura como equações de Henderson e com as mesmas são obtidas as soluções para os efeitos fixos  $\hat{\beta}$  e a predição para os efeitos aleatórios  $\hat{u}$ .

As soluções das equações são dadas por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \quad (2.21)$$

Henderson, Kempthorne e Searle (1959) provaram que  $\hat{\beta}$ , dado pela equação (2.21), coincide com a solução obtida pelo método de mínimos quadrados generalizados, cuja expressão é

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y. \quad (2.22)$$

E a predição de  $\hat{u}$ , segundo Henderson (1963), é dada por:

$$\hat{u} = GZV^{-1}(y - X\hat{\beta}) \quad (2.23)$$

em que

$$V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}.$$

Segundo Searle (1971), há desvantagem em não utilizar esta opção, pois o cálculo de  $V^{-1}$  pode se tornar difícil, uma vez que a dimensão de  $V$  é igual ao número de observações que, muitas das vezes, na área de melhoramento genético, por exemplo, chega a ser muito alta. Sendo  $G$  e  $R$  conhecidas,

$\hat{\beta}$  é o melhor estimador linear não viesado (BLUE) e  $\hat{u}$  é o melhor preditor linear não viesado (BLUP) (LITTELL et al., 2006).

#### 2.4.2 Estimação dos parâmetros das matrizes de covariâncias $R$ e $G$

Para prever e estimar os parâmetros da equação (2.21) é necessário que se conheça a estrutura de  $V$ , porém, a mesma está condicionada às estruturas de  $R$  e  $G$ , que são desconhecidas na maioria das vezes. Sendo assim, para se conhecer  $V$ , é necessário estimar os parâmetros de  $R$  e  $G$ .

A estimação de  $G$  e  $R$  pode ser realizada por vários métodos, dentre os quais destacam-se os métodos de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML) (SEARLE; CASELLA; MCCULLOCH, 1992).

O método de máxima verossimilhança consiste em obter a função de verossimilhança das observações e maximizá-la em função dos parâmetros de  $R$ ,  $G$  e dos efeitos fixos, assumindo, neste caso, a premissa de que  $y$  do modelo (2.20) apresente distribuição multinormal. Já os estimadores de máxima verossimilhança restrita são obtidos maximizando-se a parte da função de verossimilhança, que é invariante ao parâmetro de locação, isto é, em termos do modelo misto, é invariante para  $X\beta$  (PERRI; IEMMA, 1999). Atualmente, em muitos trabalhos, tem-se preferido utilizar REML, pois mantém as propriedades do ML, é não viesado e permite também imposição de não negatividade.

Pinheiro e Bates (2000) destacam os métodos de máxima verossimilhança e máxima verossimilhança restrita em relação a outros métodos.

Freitas (2007) apresenta os *log* das funções de verossimilhança e verossimilhança restrita:

1º) Log da função de verossimilhança obtida pelo método ML:

$$l(G, R) = -\frac{1}{2} \log|V| - \frac{n}{2} \log (y - X(X'V^{-1}X)X'V^{-1}y)' V^{-1} (y - X(X'V^{-1}X)X'V^{-1}y) - \frac{n}{2} [1 + \log(2\pi/n)] \quad (2.24)$$

2º) Log da função de verossimilhança obtida pelo método REML:

$$l(G, R) = -\frac{1}{2} \log|V| - \frac{1}{2} \log|X'V^{-1}X| - \frac{n-p}{2} \log (y - X(X'V^{-1}X)X'V^{-1}y)' V^{-1} (y - X(X'V^{-1}X)X'V^{-1}y) - \frac{n-p}{2} \left[ 1 + \log\left(\frac{2\pi}{n-p}\right) \right], \quad (2.25)$$

em que  $p$  é o posto da matriz  $X$ .

Em ambos os métodos, são geradas equações cujas soluções fornecem estimativas para os parâmetros de  $G$  e  $R$ . Essas estimativas são obtidas por processos iterativos, uma vez que não é possível a obtenção de expressões analíticas para os parâmetros. Portanto, são utilizados algoritmos tais como o Newton-Raphson, dentre outros. O processo é repetido até que o critério de convergência seja atendido.

Quando se substitui em (2.21), as matrizes de covariâncias estimadas  $\hat{G}$  e  $\hat{R}$ , de  $G$  e  $R$ , respectivamente, o sistema de equações passa a ser:

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'\hat{R}^{-1}y \\ Z'\hat{R}^{-1}y \end{bmatrix}. \quad (2.26)$$

As siglas BLUE e BLUP já não se aplicam em todos os aspectos às equações de modelos mistos. O estimador de  $\beta$  e o preditor de  $u$  passam a ser:

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1} X'\hat{V}^{-1}y \quad (2.27)$$

$$\hat{u} = GZ\hat{V}^{-1}(y - X\hat{\beta}). \quad (2.28)$$

Como  $\beta$  e  $u$  foram obtidos a partir de estimadores de  $\hat{G}$  e  $\hat{R}$ , uma vez que  $G$  e  $R$  são desconhecidos,  $\hat{\beta}$  e  $\hat{u}$  passam a ser denominados BLUE e BLUP empíricos, obtendo-se EBLUE e EBLUP, respectivamente (LITTELL et al., 2006). Portanto, um problema relacionado com a estimação dos parâmetros das matrizes de covariâncias consiste no fato de optar pelo método mais

conveniente, pois há vários métodos para a obtenção das estimativas dos mesmos, bem como a seleção das estruturas mais adequadas. Sendo assim, a opção pelo método mais adequado e a seleção das estruturas de covariâncias mais condizentes a estrutura dos dados tornam-se de suma importância para que  $\hat{\beta}$  e  $\hat{u}$  sejam os mais adequados possíveis.

A escolha da estrutura de covariância atinge diretamente as estimativas de efeitos fixos e erros padrões, diagnósticos e inferências. Sendo assim, esta escolha depende do perfil da estrutura dos dados, bem como de programas computacionais disponíveis.

Com a utilização da metodologia de modelos lineares mistos é possível considerar formas específicas para a matriz de covariância que representem a variabilidade real dos dados da forma mais adequada possível. Boeck, Naberezny e Tavares (2011) apresentam algumas das possíveis estruturas pré-determinadas.

A seguir são apresentadas algumas das estruturas de  $G$  e  $R$  que se encontram implementadas no SAS (LITTELL et al., 2006), considerando, por exemplo,  $n_i = 4$  ocasiões de medidas repetidas.

### 2.4.3 Estruturas das matrizes de covariâncias

1) Componente da variância (VC)

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

Esta estrutura impõe variâncias iguais em todas as ocasiões de medidas e observações independentes e tem um único parâmetro (CECON et al., 2008).

2) Simetria composta (CS)

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

Impõe variâncias iguais em todas as ocasiões e mesma covariância entre medidas realizadas em ocasiões distintas (CECON et al., 2008). Tem dois parâmetros.

3) Sem estrutura (UN)

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

Tem a particularidade de permitir que as variâncias e as covariâncias possam ser distintas para cada uma das  $n_i$  ocasiões consideradas. Ela é útil quando não há o padrão de variâncias e covariâncias. Possui  $n_i(n_i + 1)/2$  parâmetros.

4) Autorregressiva de primeira ordem AR(1)

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

Esta matriz representa que a correlação decresce à medida que as distâncias entre as medições vão aumentando. As variâncias entre todas as ocasiões são iguais (CECON et al., 2008).

5) Autorregressiva de primeira ordem médias móveis ARMA(1,1)

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

É a estrutura de séries temporais com parâmetro autorregressivo  $\rho$ , de componentes de médias móveis  $\gamma$ , com  $\sigma^2$  sendo a variância residual (BOECK; NABEREZNY; TAVARES, 2011). Possui três parâmetros. Essa matriz é uma extensão da AR(1) com a inclusão do parâmetro  $\gamma$ , no sentido de flexibilizar a modelagem das correlações.

#### 6) Toeplitz (TOEP)

Matriz caracterizada pela propriedade de possuir variâncias iguais e covariâncias desiguais para ocasiões não adjacentes (BOECK; NABEREZNY; TAVARES, 2011). É uma estrutura utilizada em séries temporais. Possui  $n_i$  parâmetros.

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

#### 7) Simetria composta heterogênea (CSH)

Nessa estrutura, as variâncias são distintas para cada elemento da diagonal principal e raiz quadrada desses parâmetros fora da diagonal principal, sendo  $\sigma_i^2$  o  $i$ -ésimo parâmetro de variância e  $\rho$  o parâmetro de correlação. Tem  $n_i + 1$  parâmetros (CECON et al., 2008).

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

### 8) Huynh-Feldt (H-F)

Matriz caracterizada por apresentar variâncias distintas nas ocasiões medidas e covariâncias obtidas pela média aritmética das variâncias subtraída da diferença entre a média das variâncias e a média das covariâncias, ou seja,  $\lambda$  (HUYNH; FELDT, 1970). Foi uma estrutura já estudada em 2.3.1. Possui  $n_i + 1$  parâmetros.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{ij} & \sigma_{ij} & \sigma_{ij} \\ \sigma_{ij} & \sigma_2^2 & \sigma_{ij} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{ij} & \sigma_3^2 & \sigma_{ij} \\ \sigma_{ij} & \sigma_{ij} & \sigma_{ij} & \sigma_4^2 \end{bmatrix}, \text{ com } \sigma_{ij} = [(\sigma_i^2 - \sigma_j^2)/2] - \lambda, \text{ para } i \neq j.$$

Segundo Littell et al. (2006), na seleção da estrutura de covariância adequada, inicialmente deve-se considerar no modelo as estruturas com maior número de parâmetro e, posteriormente, buscar estruturas mais parcimoniosas, que envolvam o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da correlação na unidade experimental.

#### 2.4.4 Seleção dos modelos

Selecionar um modelo adequado no que se refere à escolha das matrizes de covariâncias não é algo trivial, pois existem muitas estruturas pré-determinadas. Uma forma a priori de direcionar e minimizar o número de estruturas inicialmente candidatas é por meio de estudos gráficos que permitam visualizar o comportamento da variável resposta ao longo do tempo, no caso de dados longitudinais. Todavia, existem vários testes para realizar tal procedimento:

#### 2.4.4.1 Teste assintótico da razão de verossimilhança

Segundo Pinheiro e Bates (2000), o teste da razão de verossimilhança é um teste estatístico utilizado para comparar o ajuste de dois modelos estimados por máxima verossimilhança, sendo o modelo nulo caso especial do outro, ou seja, o modelo alternativo. O teste é baseado no logaritmo natural da razão de duas verossimilhanças do modelo mais simples  $L_1$  com menor número de parâmetros e o modelo mais completo  $L_2$ . A estatística do teste é:

$$-2\ln(L_1/L_2) = 2[\ln(L_2) - \ln(L_1)] \sim \chi_r^2$$

Se  $k_i$  é o número de parâmetros a serem estimados no modelo  $i$ , então, assintoticamente a distribuição da estatística de  $2\ln(L_1/L_2)$ , sob hipótese nula que o modelo mais simples é adequado, segue uma distribuição de qui-quadrado com  $r = K_2 - K_1$  graus de liberdade.

Segundo os mesmos autores, este teste pode ser usado para modelos ajustados pelo método de máxima verossimilhança restrita, porém, ambos os modelos devem ter sido ajustados por esse método e ter a mesma estrutura para os efeitos fixos.

É um critério que deve ser utilizado para comparar dois modelos de cada vez, sendo um modelo caso especial do outro.

#### 2.4.4.2 Critérios de informação

Os critérios de informação são uma alternativa ao teste da razão de verossimilhança e são valores para os logaritmos das funções de verossimilhanças do modelo, que estão em função do número de parâmetros e do número de observações. Nesses casos, é selecionado o modelo que apresentar menor valor, sendo penalizado o modelo com maior número de parâmetros.

O critério de informação de Akaike (AIC) e o de Schwarz (BIC) são métodos de seleções de modelos que podem ser utilizados para comparar modelos aninhados e não aninhados, ou seja, quando um é, ou não, caso especial do outro. Esses critérios não utilizam o conceito de significância estatística na sua aplicação, pois não há formulação de uma hipótese, não sendo necessário decidir sobre um valor- $p$  limite que se considera estatisticamente significativo (MOTULSKY; CHRISTOPOULOS, 2003). O método permite determinar qual modelo é mais provável que esteja correto, por meio da quantificação do quanto mais provável é o modelo.

Conforme Littell et al. (2006), os critérios de informação utilizando REML podem ser obtidos da seguinte forma:

$$AIC = -2\ln(\hat{\theta}) + 2d \quad (2.29)$$

$$BIC = -2\ln(\hat{\theta}) + d \ln(n - \text{posto}(X)) \quad (2.30)$$

em que  $\ln(\hat{\theta})$  é o logaritmo natural da função de verossimilhança,  $d$  é o número de parâmetros de covariância e  $n$  é o número total de observações. Ainda segundo os autores, não há uma forma clara de obter esses valores. Suas expressões podem variar, principalmente com o uso de diferentes softwares. Será considerado o modelo mais plausível aquele que apresentar o menor valor, isto segundo ambos os critérios.

Littell et al. (2006) e Pinheiro e Bates (2000) alertam para o fato de que os modelos ajustados por REML não são comparáveis pelos critérios anteriores quando possuírem diferentes estruturas de efeitos fixos. Entretanto, quando são ajustados por ML, não ocorre a necessidade de possuírem idêntica estrutura de efeitos fixos, para a comparação de modelos.

### 2.4.5 Inferência sobre os efeitos fixos e aleatórios

Os procedimentos para realizar inferências acerca dos parâmetros de efeitos fixos e aleatórios são descritos por Littell et al. (2006). Para realizar inferências nos parâmetros de efeitos fixos  $\beta$  e aleatórios  $u$  no modelo misto, são consideradas funções lineares estimáveis da seguinte forma:

$$\omega = L \begin{bmatrix} \beta \\ u \end{bmatrix}$$

Funções dessa forma são chamadas de funções previsíveis se a parte fixa  $\beta$  da matriz  $L$  atender à exigência de estimabilidade. Como a inferência sobre os efeitos fixos é o foco principal, a parte de  $u$  associada a  $L$  assume valores nulos.

Inferências estatísticas podem ser obtidas por meio da formulação da seguinte hipótese nula:

$$H_0 = L \begin{bmatrix} \beta \\ u \end{bmatrix} = 0$$

Quando  $L$  é composta por uma única linha, uma estatística  $t$  geral pode ser construída, como segue abaixo:

$$t = \frac{L \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}}{\sqrt{L\hat{C}L'}} \quad (2.31)$$

em que

$$\hat{C} = \begin{bmatrix} X'\hat{R}^{-1}X & X'(\hat{R}^{-1})Z \\ Z'\hat{R}^{-1}X & Z'R^{-1}\hat{Z} + G^{-1} \end{bmatrix}^{-1}. \quad (2.32)$$

Assumindo as pressuposições de  $u$  e  $e$  do modelo (2.20) serem distribuídos normalmente,  $t$  geralmente terá distribuição aproximada.

Denotando  $\nu$  como os graus de liberdade associado a  $\omega$ , o intervalo de confiança de  $\omega$  será o seguinte:

$$L \begin{bmatrix} \beta \\ u \end{bmatrix} \pm t_{(\nu, \alpha/2)} \sqrt{L\hat{C}L'}$$

sendo  $t_{(v, \alpha/2)}$  percentil  $(1 - \alpha/2)100\%$  da distribuição  $t_v$ .

Ocorre que o posto de  $L$  pode ser maior que 1 e, neste caso, deve-se considerar a estatística  $F$ :

$$F = \frac{\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}' L'(L\hat{C}L')^{-1}L \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}}{\text{posto}(L)} \quad (2.33)$$

De forma semelhante à distribuição  $t$ , a distribuição  $F$ , em geral, tem distribuição aproximada com número de graus de liberdade do numerador igual ao  $\text{posto}(L)$  e  $\hat{v}$  como número de graus de liberdade do denominador. Portanto, as estatísticas  $t$  e  $F$  permitem fazer inferências sobre os efeitos fixos estimados para o modelo covariância selecionado.

## 2.5 Análise de medidas repetidas com ingestão de café

O conjunto de dados provenientes do experimento que foi estudado nesta dissertação foi também utilizado por Barbosa (2007) e Silva (2009), em suas respectivas dissertações de mestrado. Silva (2009) realizou apenas ampla análise exploratória do conjunto de dados e a análise univariada de perfil, utilizando o esquema de parcela subdividida no tempo, não abordando outra técnica ou, até mesmo, a análise univariada de perfil corrigida. Barbosa (2007) realizou apenas a análise univariada de perfil, entretanto, abordou vários outros aspectos do experimento de interesse da área de farmácia. Também não enfatizou outra metodologia estatística, assim como não abordou o fato de que os pesos de um mesmo animal (rato) poderiam estar, ou não, com alguma estrutura de dependência, devido ao fato de terem sido obtidos ao longo de quinze semanas.

### 3 MATERIAL E MÉTODOS

Nesta seção são descritos a metodologia e os dados utilizados neste trabalho.

#### 3.1 Material

Os dados utilizados para exemplificação e comparação das técnicas de análise de dados com medidas repetidas são provenientes de experimento realizado por Barbosa (2007) em sua dissertação de mestrado e também usados por Silva (2009). Foram coletados pesos de ratos submetidos a diferentes dietas (tratamentos) semanalmente, totalizando 15 semanas, tendo sido aferidos da quarta até a décima oitava semana de vida dos referidos animais. Melhores detalhes do experimento podem ser vistos em Barbosa (2007) e Silva (2009).

Os tratamentos foram constituídos de dietas, sendo três sem extrato de café e três com extrato de café, os quais foram descritos por Barbosa (2007) como:

- . (NORM) grupo normonutrido controle, recebeu ração de biotério para ratos (Labina.) enriquecida com ração para cães (Bonzo);

- . (NCAF) grupo normonutrido café, foi alimentado com dieta do grupo normonutrido controle acrescida de extrato aquoso de café;

- . (DESN) grupo desnutrido controle, foi alimentado com dieta correspondente a 60% da ração consumida pelo grupo normonutrido controle;

- . (DCAF) grupo desnutrido café, recebeu dieta do grupo desnutrido controle acrescida de extrato aquoso de café;

- . (DREC) grupo desnutrido com condição alimentar reconstituída controle, foi alimentado com dieta correspondente a 60% da ração consumida pelo grupo normonutrido controle e, após 60 dias, foi alimentado *ad libitum*;

. (DRECAF) grupo desnutrido com condição alimentar reconstituída café, recebeu dieta do grupo desnutrido acrescida de extrato aquoso de café e, após 60 dias, alimentado *ad libitum*.

Todos os grupos receberam água *ad libitum* e foram pesados semanalmente, durante 15 semanas.

### **3.1.1 Modelo experimental**

Ratos da raça Wistar, da colônia do Laboratório de Nutrição Experimental da Faculdade de Farmácia da UFMG, foram acasalados e, por volta do 18º dia de prenhez, as fêmeas foram transferidas para gaiolas individuais e divididas aleatoriamente em seis grupos, como descrito no tópico anterior. Todas as dietas experimentais foram oferecidas a partir do nascimento dos filhotes. O número de filhotes por fêmea foi fixado em oito, tendo, posteriormente, somente os machos sido mantidos. Os animais tinham livre acesso à água e eram mantidos em sistema claro/escuro (12/12 horas).

Após o desmame realizado aos 21 dias de vida, filhotes de ratos da raça Wistar foram separados, mantendo-se a mesma dieta oferecida para a respectiva fêmea matriz, sendo avaliados semanalmente, durante 15 semanas.

## **3.2 Métodos**

### **I) Análise exploratória**

Primeiramente, foi realizada a análise exploratória por meio de procedimentos gráficos. Os procedimentos gráficos utilizados foram plotagem dos dados reais para cada dieta alimentar, estudo da dispersão dos dados por meio da construção de gráficos do tipo boxplot para cada dieta alimentar,

unidade experimental e semanas, bem como a plotagem dos pesos médios dos animais (ratos) em cada dieta alimentar.

O intuito principal desta análise gráfica foi o de estudar o comportamento do peso dos animais em relação aos grupos de dietas, tempo (semanas) e interação entre eles. Também se verificou a presença de valores atípicos, bem como da correlação entre os pesos dos animais dos diferentes tempos, com a finalidade de verificar possível relação entre os pesos do mesmo animal em idades diferentes.

## II) Análise univariada de perfil

Nesta análise, foi utilizado o esquema de parcela subdividida no tempo, em que o fator da parcela (entre indivíduos) é aquele cujos níveis permanecem constantes, ao passo que o fator da subparcela (intraindivíduos) e a interação entre eles variam. Portanto, as dietas alimentares foram consideradas como o fator da parcela e semanas e a interação semanas×dietas alocadas na subparcela. O modelo utilizado foi

$$y_{ijk} = \mu + \tau_i + \delta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + e_{ijk}$$

em que  $y_{ijk}$  é o valor observado para o peso do animal na  $j$ -ésima unidade experimental (animal) submetida à  $i$ -ésima dieta na  $k$ -ésima semana;  $\mu$  é uma constante comum a todos os pesos;  $\tau_i$  é o efeito da  $i$ -ésima dieta;  $\delta_{j(i)}$  representa o efeito aleatório da  $j$ -ésima unidade experimental (animal) dentro da  $i$ -ésima dieta;  $\gamma_k$  é o efeito da  $k$ -ésima semana;  $(\tau\gamma)_{ik}$  é o efeito da interação entre a  $i$ -ésima dieta com a  $k$ -ésima semana e  $e_{ijk}$  é o erro aleatório associado aos pesos avaliados na  $k$ -ésima semana para a  $i$ -ésima dieta e  $j$ -ésima unidade experimental (animal), supostos homocedásticos, independentes e normalmente distribuídos, assim como o efeito aleatório da  $j$ -ésima unidade experimental (animal) dentro da  $i$ -ésima dieta, para  $i = 1, \dots, 6$ ,  $j = 1, \dots, 60$  e  $k = 1, \dots, 15$ .

Para que haja validação dos resultados, uma pressuposição dessa técnica é que a matriz de covariâncias atenda à condição de Huynh e Feldt (1970). Para verificar essa pressuposição, foi utilizado o teste de esfericidade de Mauchly. Se a matriz de covariâncias não atender à condição de Huynh e Feldt (1970), pelo teste Mauchly, a primeira alternativa considerada foi realizar o ajuste dos graus de liberdade do teste F das fontes de variação intraindivíduo pelas correções de Geisser e Greenhouse (1958) e também Huynh e Feldt (1976). A segunda alternativa da análise foi a realização da análise multivariada de perfil.

### III) Análise multivariada de perfil

A análise multivariada de perfil é uma alternativa natural a análise dos dados da variável peso, pois não requer estrutura específica de covariância para representar a correlação possivelmente existente na parcela experimental, consequência do fato de os dados serem oriundos de períodos temporais. Ao utilizar essa técnica, foi considerado o modelo

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$$

em que  $y_{ijk}$  é o valor observado do peso do animal na  $i$ -ésima dieta;  $j$ -ésima semana na  $k$ -ésima repetição;  $\mu$  é uma constante comum a todos os pesos;  $\tau_i$  é o efeito da  $i$ -ésima dieta;  $\gamma_j$  é o efeito da  $j$ -ésima semana;  $(\tau\gamma)_{ij}$  é o efeito da interação entre a  $i$ -ésima dieta com a  $j$ -ésima semana;  $e_{ijk}$  é o erro aleatório associado ao peso avaliado na  $i$ -ésima dieta  $j$ -ésima semana e  $k$ -ésima repetição, supostos homocedásticos, independentes e distribuídos como normal multivariada, para  $i = 1, \dots, 6$  dietas,  $j = 1, \dots, 15$  semanas e  $k = 1, \dots, 10$  repetição.

Esse modelo diferencia-se do modelo considerado na análise univariada, pois não foi considerado o erro da parcela.

#### IV) Modelo linear misto

Por fim, com ênfase no ajuste de curvas polinomiais de crescimento para cada dieta, é utilizada a abordagem do modelo linear misto. Ao utilizar a técnica de análise de dados provenientes de medidas repetidas no tempo via modelo linear misto, seguem-se às seguintes etapas: seleção dos efeitos fixos, que está relacionado com os perfis médios de respostas; identificação dos efeitos aleatórios, que estão relacionados com os indivíduos (parcela experimental) e escolha da melhor estrutura para a matriz de covariância.

Os efeitos fixos considerados, assim como nas análises anteriores, foram dietas, semanas e interação de semanas  $\times$  dietas. Para a inclusão dos efeitos aleatórios, foi sugerida a análise de regressão polinomial para cada parcela experimental (combinação de dieta e repetição), no qual foram selecionados os efeitos aleatórios a partir do estudo da significância dos parâmetros dos respectivos polinômios pelo teste *t*-Student. Já o grau do polinômio para representar a tendência do perfil médio do peso ao longo das semanas em cada dieta foi sugerido pela análise de gráficos de perfis médios dos pesos construídos na análise exploratória e confirmado pela análise de regressão realizada para os perfis médios dos pesos para cada dieta.

Considerando os efeitos fixos anteriores e que o comportamento dos pesos médios dos animais relacionado a cada dieta possa ser explicado por curvas polinomiais quadráticas, por exemplo, e considerando o intercepto, o coeficiente linear e o coeficiente quadrático das curvas individuais como aleatórios, o modelo linear misto escrito na forma geral, como em Faraway (2006) e Laird e Ware (1982), para cada unidade experimental *i*, foi

$$y_i = X_i\beta + Z_iu_i + e_i$$

em que  $y_i$  é o vetor de observações de dimensão  $t \times 1$  referente à unidade experimental *i* e tempo *t*, dado em semanas;  $X_i$  é a matriz, de dimensão  $t \times p$ , relativa aos parâmetros de efeitos fixos (dietas, semanas, interação de semanas  $\times$

dietas e interação de dietas  $\times$  semanas<sup>2</sup>, sendo esse último efeito incluído após considerar um ajuste quadrático) ligando  $\beta$  a  $y_i$ ;  $\beta$  é um vetor, de dimensão  $p \times 1$ , desconhecido, formado pelos efeitos fixos;  $Z_i$  é a matriz, de dimensão  $t \times q$ , relativa aos efeitos aleatórios (intercepto, coeficiente linear e o coeficiente quadrático das curvas individuais  $i$ , sendo cada componente expresso por  $u_{0i}, u_{1i}$  e  $u_{2i}$  respectivamente) ligando  $u_i$  a  $y_i$ ;  $u_i$  é um vetor de efeitos aleatórios ( $u_{0i}, u_{1i}$  e  $u_{2i}$ ) desconhecidos de dimensão  $q \times 1$  e  $e_i$  vetor de erros aleatórios não observáveis, de dimensão  $t \times 1$ .

Na descrição anterior, o vetor de erros aleatórios  $e_i$  associado à unidade experimental  $i$  é distribuído como  $N(0, R)$ , ou seja, normal com média 0 e matriz de covariância  $R$ , sendo a matriz  $R$  positiva definida de dimensão  $t \times t$ . Também, o vetor de erros aleatórios  $u_i = (u_{0i}, u_{1i}, u_{2i})$  distribuído como  $N(0, G)$ , sendo a matriz  $G$  positiva definida de dimensão  $q \times q$ .

Procedeu-se a escolha da melhor estrutura para a matriz de covariância  $G$  e  $R$ , utilizando-se os critérios de informação de Akaike e bayesiano de Schwarz, conforme Pinheiro e Bates (2000). Casos esses testes selecionem estruturas distintas, será utilizado o teste assintótico da razão de verossimilhança, de acordo com Littell et al. (2006), para selecionar, entre as duas estruturas selecionadas, a mais adequada.

A estimação dos parâmetros das matrizes de covariâncias selecionadas foi realizada pelo método de máxima verossimilhança restrita.

A estatística  $F$ , citada no tópico 2.4.5, foi utilizada para testar a significância dos efeitos fixos do modelo, assim como a estatística  $t$  foi utilizada para comparar contrastes formados pelas dietas alimentares e que foram classificadas de interesse prático. Essa mesma estatística também foi utilizada para testar a significância dos parâmetros das curvas polinomiais, construídas para as dietas alimentares.

O diagnóstico da normalidade dos erros do modelo foi realizado por intermédio de gráficos construídos para essa finalidade e confirmado pela aplicação do teste de Shapiro-Wilk.

## V) Softwares utilizados

Para a realização da análise exploratória foi utilizado o programa R (R DEVELOPMENT CORE TEAM, 2011). Foram utilizadas bibliotecas (pacotes) distintas, pois foram construídos gráficos que necessitaram de bibliotecas específicas. Para a realização da análise univariada e testes, foi utilizado o mesmo programa. Entretanto, para a realização da análise univariada de perfil e testes, bem como para a realização do teste de esfericidade de Mauchly, foram implementadas rotinas específicas.

A análise multivariada foi realizada utilizando-se o proc GLM do programa SAS® (SYSTEM ANALYSIS STATISTICAL INSTITUTE - SAS INSTITUTE, 2004). O proc GLM requer que os dados sejam completos nas unidades experimentais, o que garante que não haja parcelas perdidas e, se houver parcelas perdidas, ele excluirá todos os dados da unidade experimental correspondente.

Utilizando o modelo linear misto no ajuste das curvas polinomiais de crescimento para as dietas alimentares, foi utilizado o proc MIXED do SAS® (SAS INSTITUTE, 2004). Para a seleção da estrutura de covariância mais adequada para  $G$  e  $R$ , foram consideradas distintas estruturas pré-existentes no proc MIXED, selecionando-se a mais plausível pelos testes citados, já implementados no tópico anterior e estimando seus parâmetros pelo método de máxima verossimilhança restrita. Essas estruturas devem ser indicadas, pois, caso contrário, o programa usa por *default* a estrutura componente de variância. Os testes que foram aplicados nos efeitos fixos do modelo, nos contrastes de

interesse e nos parâmetros das curvas médias estimadas também já estão implementados no proc MIXED.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise exploratória

Analisando-se os perfis médios de crescimento dos pesos dos animais, (Figura 1), percebe-se que não há nenhum valor que extrapola os valores reais observados, indicando que não há diagnóstico de qualquer valor anormal que se distinga dos valores observados dos pesos dos animais e, por conseguinte, ausência de valores atípicos. Outra observação que merece destaque é a tendência de crescimento quadrático dos pesos dos animais, no desencadear do tempo. Portanto, o polinômio de segundo grau se mostra candidato prévio a representar as curvas médias de crescimento, ainda a serem estimadas em tópicos posteriores.

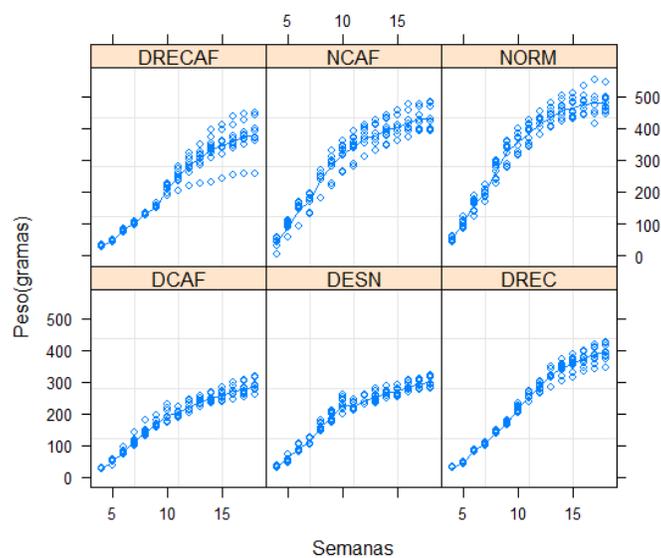


Figura 1 Representação gráfica do perfil individual do peso dos animais, em gramas, em função dos tempos de avaliação (semanas), para as diferentes dietas alimentares

Analisando-se o perfil médio de respostas dos pesos dos animais em todas as unidades experimentais (Figura 2), a suposição de variância comum entre elas parece não existir, pela evidência de diferenciação de variâncias entre muitas das unidades experimentais. Isso já era esperado, por ser uma característica comum em dados mensurados ao longo do tempo e também ressaltar que esse fato deve ser levado em consideração em análise estatística mais completa, como a análise multivariada.

Já na Figura 3, observa-se a evolução do peso dos animais ao longo das 15 semanas. A evolução dos pesos mostra, de forma nítida, a diferença no desenvolvimento do peso dos animais a cada semana, algo que tende a indicar que seja significativo o efeito do tempo no experimento. Algo já esperado, pela própria natureza dos valores, porque é natural que os animais adquiram peso, pois estão em fase de crescimento.

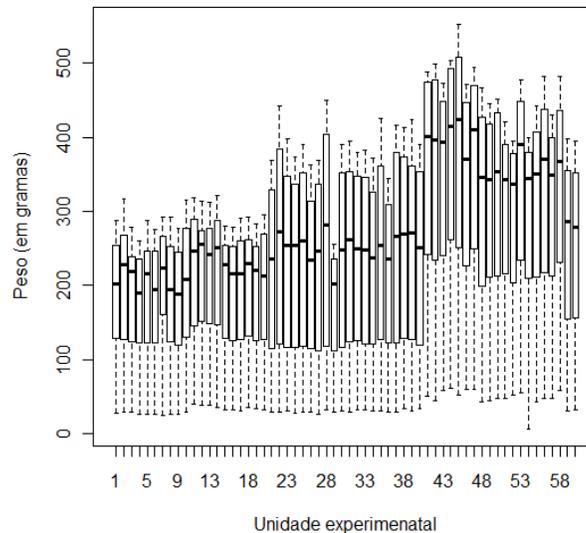


Figura 2 Representação gráfica do peso dos animais, em gramas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio em cada unidade experimental

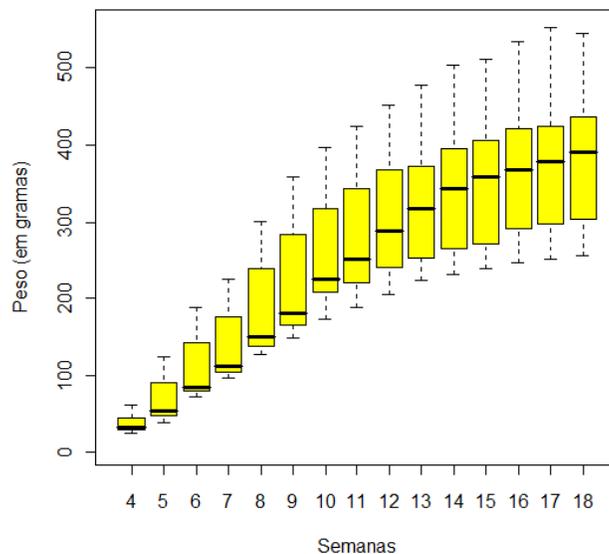


Figura 3 Representação gráfica do peso dos animais, em grammas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio do peso dos animais em cada semana avaliada

Na Figura 4 há evidência que há diferença do ganho de peso entre os animais, podendo indicar diferença significativa entre o peso médio dos animais que receberam as dietas DESN com DREC e NORM. Nessas dietas inexistiu o extrato aquoso de café. Logo, o uso de diferentes dietas tende a influenciar os pesos dos animais. Os respectivos grupos de animais que receberam as dietas DCAF e DRECAF se diferenciaram na dieta apenas pela inclusão do extrato aquoso de café e, de forma semelhante, os grupos DREC e DRECAF, bem como os grupos NCAF e NORM, respectivamente, ou seja, os grupos que receberam as dietas representadas pela cor marrom tiveram o extrato aquoso de café incluído na dieta. Entre os respectivos três grupos de animais que receberam dietas com ausência do extrato aquoso de café é nítida a influência das dietas no ganho de peso do animal, ou seja, vê-se maior ganho de peso dos animais que

receberam a dieta NORM em relação a DREC e este grupo sobre DESN. Já com a utilização do café, houve, no geral, tendência de diminuir o ganho de peso dos animais, analisando, sobretudo, os pares grupos que receberam as mesmas dietas alimentares, diferenciando-se apenas pela inclusão do extrato aquoso.

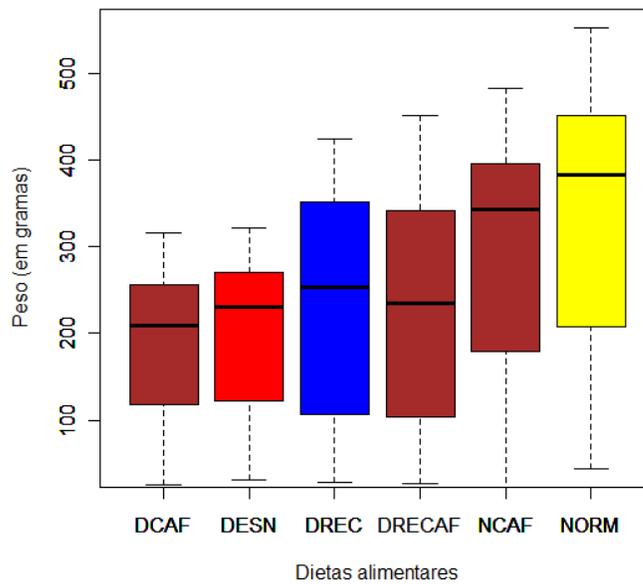


Figura 4 Representação gráfica do peso dos animais, em gramas, mostrando os limites inferior e superior da amplitude de variação, primeiro e terceiro quartil e valor médio em cada dieta alimentar

Na Figura 5 percebe-se crescimento contínuo dos animais ao longo das semanas avaliadas, para as diferentes dietas. Nota-se efeito diferenciado entre as dietas, analisando, por exemplo, as dietas DESN e DREC, assim como DCAF e DRECAF. Também pode ser observado que a inclusão do extrato aquoso de café proporcionou menor crescimento, como se nota pelo comportamento das dietas NORM e NCAF, principalmente após a 10ª semana.

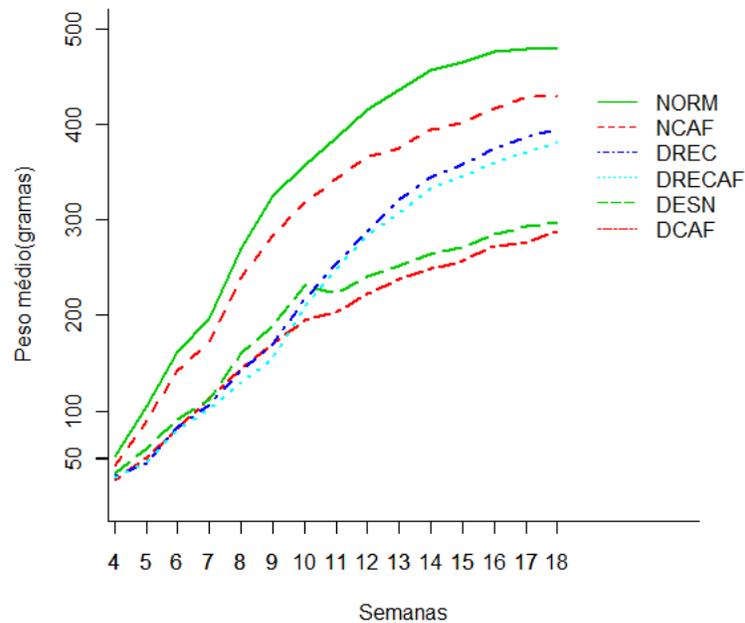


Figura 5 Representação gráfica do peso médio dos animais em relação ao período de avaliação (semanas), para as seis dietas utilizadas

#### 4.2 Análise univariada de perfil

Os resultados da análise de variância univariada (Tabela 2) do peso dos animais indicam que existem diferenças significativas entre as dietas, semanas e interação semanas  $\times$  dietas. Sendo assim, a hipótese de perfis coincidentes que é testada para o fator entre dietas, ou fator da parcela, foi rejeitada pelo teste F, a 1% de probabilidade, o que indica que as dietas têm efeitos distintos sobre o peso médio dos animais. Para o teste da hipótese de perfis horizontais presentes na fonte de variação entre semanas, que testa o efeito de semanas, também foi rejeitado pelo teste F, a 1% de probabilidade, indicando que a evolução do tempo afeta o peso médio dos referidos grupos de animais.

Também para o teste da hipótese de perfis paralelos presentes na fonte e variação intraindivíduo, que testa o efeito da interação semanas  $\times$  dietas, foi rejeitada pelo teste F, a 1% de probabilidade, indicando que a interação de dietas com semanas age de forma dependente sobre o peso médio dos animais.

Entretanto, para que os resultados apresentados anteriormente sejam confiáveis, é necessário que seja avaliado o resultado do teste de Mauchly apresentado na Tabela 3, no próximo tópico.

Tabela 2 Resultado da análise de variância univariada da variável peso por animal (grama), considerando o delineamento inteiramente casualizado e com o fator grupos de dietas nas parcelas e os períodos de avaliação nas subparcelas.

Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F	Valor - p
Dietas	5	2563527	512705	88,66	<0,001
Erro(a)	54	312268	5782		
Semanas	14	11402026	814430	4795,73	<0,001
Sem. $\times$ Dietas	70	563291	8047	47,38	<0,001
Erro(b)	756	128387	169,9		
Coeficiente de variação = 13,07%					

#### 4.2.1 Teste de Mauchly

Pelo resultado apresentado na Tabela 3, constata-se que o teste de Mauchly rejeitou a hipótese de esfericidade, o que implica que a matriz de covariâncias do erro experimental não atende à condição proposta por Huynh-Feldt. Portanto, os resultados presentes na subparcela podem não ser confiáveis

e uma alternativa é proceder a uma correção do número do grau de liberdade das fontes de variação da subparcela.

Tabela 3 Resultado do teste de esfericidade de Mauchly.

Graus de liberdade $f$	Critério de Mauchly ( W)	Estatística de Qui - quadrado $\chi^2$	Valor - p
104	1,12E-11	1239,1275	<0,0001

É comum, em ciências agrárias, que o teste de Mauchly rejeite a hipótese nula de esfericidade da matriz de covariância. Na pesquisa realizada por Freitas et al. (2007), por exemplo, em um experimento em que se avaliou o efeito de distintos genótipos de cana-de-açúcar, houve rejeição da hipótese nula de esfericidade da matriz de covariância da análise de variância considerada.

#### 4.2.2 Análise univariada de perfil corrigida para as fontes de variação semanas e interação de dietas com semanas

Como apresentado na Tabela 4, mesmo com a utilização da correção dos números de graus de liberdade do teste F pelas correções de Geisser-Greenhouse e Huynh-Feldt (1958), todos os resultados permaneceram inalterados, quando comparados aos da análise univariada de perfil, indicando que as hipóteses de perfis horizontais (efeito nulo de semanas) e a hipótese de perfis paralelos (ausência de interação semanas×dietas) foram rejeitadas.

Os resultados obtidos pela análise de variância univariada corrigida do peso dos animais confirmam a significância de semanas e interação semanas ×

dietas, diagnosticada por Barbosa (2007) e Silva (2009), ao realizarem a análise de variância univariada em suas pesquisas.

Tabela 4 Resultado da análise univariada de perfil corrigida em um delineamento inteiramente casualizado para os efeitos intraindivíduos do peso do animal submetido a 6 dietas, aferido no período de 15 semanas, a partir da 4<sup>a</sup> semana de vida.

Teste para os fatores intraindivíduos (subparcela)					
Fontes de variação	Graus de liberdade	Soma de quadrados	Quadrad o médio	F	Valor - p
Semanas	14	11402026	814430	4795,73	<0,001
Sem.×Dietas	70	563291	8047	47,38	<0,001
Erro(b)	756	128387	169,9		
Ajuste do teste F com correções dos graus de liberdade					
Fontes de variação	G-G	H-F			
Semanas	<0,001	<0,001			
Sem.×Grupos	<0,001	<0,001			
G-G= Grenhouse-Geisser $\hat{\epsilon} = 0,1265$ e H-F= Huynh – Feldt $\tilde{\epsilon} = 0,1426$					

O valor  $\hat{\epsilon} = 0,1265$  é a estimativa da correção proposta por Geisser e Greenhouse (1958) e  $\tilde{\epsilon} = 0,1426$  é a estimativa da correção proposta por Huynh e Feldt (1976), que realizaram o ajuste do número de grau de liberdade das fontes de variação da subparcela.

Em experimento avaliado por Amador (2010), que estudou diferentes densidades de cultivo da planta mamona, também a correção dos graus de liberdade não modificou os resultados da análise de variância realizada.

Portanto, como não houve alteração dos resultados provocada pelo uso das correções dos graus de liberdade do teste F, referente às fontes de variações intraindivíduos (subparcela), a metodologia da análise multivariada de perfil pode ser considerada como metodologia alternativa mais coerente, por não apresentar restrições quanto à matriz de covariâncias do erro.

### **4.3 Análise multivariada de perfil**

Nos resultados é impressa a matriz  $R$ , que é a matriz das correlações estimadas a partir da matriz  $E$ , entre os valores dos pesos dos animais, no decorrer das 15 semanas avaliadas.

Avaliando-se os valores das correlações entre as semanas constantes da matriz  $R$ , constata-se que, em sua maioria, eles referendam a hipótese aventada por Littell et al. (2006), bem como Xavier e Dias (2001), de que medidas tomadas em tempos mais próximos são mais fortemente correlacionadas.

$$R = \begin{bmatrix} \text{Se4} & \text{Se5} & \text{Se6} & \text{Se7} & \text{Se8} & \text{Se9} & \text{Se10} & \text{Sm11} & \text{Sm12} & \text{Se13} & \text{Se14} & \text{Se15} & \text{Se16} & \text{Se17} & \text{Se18} \\ \text{Se4} & 1 & 0,88 & 0,85 & 0,74 & 0,75 & 0,43 & 0,45 & 0,65 & 0,60 & 0,40 & -0,21 & 0,40 & 0,42 & 0,37 & 0,17 \\ \text{Se5} & 0,88 & 1 & 0,91 & 0,80 & 0,81 & 0,50 & 0,53 & 0,72 & 0,66 & -0,07 & -0,08 & 0,48 & 0,49 & 0,37 & 0,18 \\ \text{Se6} & 0,85 & 0,91 & 1 & 0,91 & 0,95 & 0,56 & 0,60 & 0,83 & 0,77 & -0,03 & -0,03 & 0,55 & 0,57 & 0,50 & 0,24 \\ \text{Se7} & 0,74 & 0,80 & 0,91 & 1 & 0,92 & 0,45 & 0,48 & 0,78 & 0,73 & -0,05 & -0,04 & 0,57 & 0,56 & 0,49 & 0,25 \\ \text{Se8} & 0,75 & 0,81 & 0,95 & 0,92 & 1 & 0,60 & 0,64 & 0,87 & 0,80 & -0,03 & -0,02 & 0,59 & 0,61 & 0,56 & 0,28 \\ \text{Se9} & 0,43 & 0,50 & 0,56 & 0,45 & 0,60 & 1 & 0,98 & 0,68 & 0,64 & 0,02 & 0,02 & 0,41 & 0,43 & 0,40 & 0,20 \\ \text{Se10} & 0,45 & 0,53 & 0,60 & 0,48 & 0,64 & 0,98 & 1 & 0,74 & 0,70 & 0,04 & 0,04 & 0,48 & 0,51 & 0,47 & 0,27 \\ \text{Se11} & 0,65 & 0,72 & 0,83 & 0,78 & 0,87 & 0,68 & 0,74 & 1 & 0,94 & 0,07 & 0,10 & 0,84 & 0,86 & 0,82 & 0,46 \\ \text{Se12} & 0,60 & 0,66 & 0,77 & 0,73 & 0,80 & 0,64 & 0,70 & 0,94 & 1 & 0,09 & 0,13 & 0,86 & 0,89 & 0,85 & 0,50 \\ \text{Se13} & -0,21 & -0,07 & -0,03 & -0,05 & -0,03 & 0,02 & 0,04 & 0,07 & 0,09 & 1 & 0,99 & 0,22 & 0,20 & 0,19 & 0,10 \\ \text{Se14} & -0,21 & -0,08 & -0,03 & -0,04 & -0,02 & 0,02 & 0,04 & 0,10 & 0,13 & 0,99 & 1 & 0,26 & 0,25 & 0,24 & 0,11 \\ \text{Se15} & 0,40 & 0,48 & 0,55 & 0,57 & 0,59 & 0,41 & 0,48 & 0,84 & 0,86 & 0,22 & 0,26 & 1 & 0,97 & 0,96 & 0,56 \\ \text{Se16} & 0,42 & 0,49 & 0,57 & 0,56 & 0,61 & 0,43 & 0,51 & 0,86 & 0,89 & 0,20 & 0,25 & 0,97 & 1 & 0,97 & 0,58 \\ \text{Se17} & 0,37 & 0,44 & 0,50 & 0,49 & 0,56 & 0,40 & 0,47 & 0,82 & 0,85 & 0,19 & 0,24 & 0,96 & 0,97 & 1 & 0,55 \\ \text{Se18} & 0,17 & 0,18 & 0,24 & 0,25 & 0,28 & 0,20 & 0,27 & 0,46 & 0,50 & 0,10 & 0,11 & 0,56 & 0,58 & 0,55 & 1 \end{bmatrix}$$

Ao usar o proc GLM, foram impressos os quatro testes multivariados Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Roy, sendo todos independentes da estrutura da matriz de covariância.

Pelos testes multivariados (Tabela 5), testou-se a hipótese nula de perfis horizontais, ou seja, que testa o efeito do fator semanas. Para a fonte de variação que foi testada, foi obtida uma matriz  $H$  de soma de quadrados e produtos, além da matriz  $E$  de soma de quadrados e produtos referentes aos erros do modelo. As estatísticas de teste foram obtidas a partir das raízes latentes da matriz  $HE^{-1}$ .

Na Tabela 5 foram usados a matriz  $H$  da soma de quadrados e produtos para o fator semanas e a matriz  $E$  soma de quadrados e produtos dos erros. Além disso, foram impressos os valores  $S$ ,  $m_1$  e  $m_2$ , utilizados para obter a aproximação com a estatística F.

Tabela 5 Resultado da análise multivariada da variável peso (grama) por animal, considerando o delineamento inteiramente casualizado e testando o efeito nulo do fator semanas.

$S = 1$ , $m_1 = 6$ e $m_2 = 19,5$					
Estatística	Valor	F	G.L. Num <sup>1</sup> .	G.L. Den <sup>2</sup> .	Valor – p
Lambda de Wilks	0,0025	1138,83	14	41	<0,001
Traço de Pillai	0,9974	1138,83	14	41	<0,001
Traço de Hotelling- Lawley	388,8680	1138,83	14	41	<0,001
Raiz máxima de Roy	388,8680	1138,83	14	41	<0,001

Pelos testes multivariados (Tabela 6), testou-se a hipótese nula de perfis paralelos que testa a interação entre dietas  $\times$  semanas. As estatísticas de teste foram obtidas a partir das raízes latentes da matriz  $HE^{-1}$ .

Na Tabela 6, assim como na Tabela 5, primeiramente, foram utilizadas a matriz  $H$  da soma de quadrados e produtos para a interação de semanas  $\times$  dietas e a matriz  $E$  soma de quadrados e produtos dos erros. Além disso, foram impressos os valores  $S$ ,  $m_1$  e  $m_2$ , utilizados para obter a aproximação com a estatística  $F$ .

Tabela 6 Resultado da análise multivariada da variável peso (grama) por animal, considerando o delineamento inteiramente casualizado e testando efeito nulo da interação de semanas  $\times$  dietas.

$S = 5$ , $m_1 = 4$ e $m_2 = 19,5$					
Estatística	Valor	F	G.L. Num <sup>1</sup> .	G.L. Den <sup>2</sup> .	Valor – p
Lambda de Wilks	0,0002	12,84	70	199,29	<0,001
Traço de Pillai	3,3496	6,52	70	225,00	<0,001
Traço de Hotelling- Lawley	45,8301	25,84	70	129,07	<0,001
Raiz máxima de Roy	30,9503	9948	14	45	<0,001

1: Graus do numerador. 2: Graus de liberdade do denominador.

Observa-se, na Tabela 5, que a hipótese de perfis horizontais que testa o efeito do fator semanas e a hipótese de perfis paralelos (Tabela 6), que testa o efeito da interação de dietas  $\times$  semanas, como já citado anteriormente, foram, ambas, rejeitadas por todos os testes, a 1% de probabilidade. Isso indica, pela hipótese de paralelismo, que pelo menos uma combinação de dietas e semanas

interage de forma dependente. Esses resultados não se distinguiram dos resultados obtidos na análise univariada de perfil, mesmo que o teste de Mauchly tenha indicado que a matriz de covariância não atendeu à condição de Huynh-Feldt.

É válido ressaltar que, nem sempre, esses testes resultam em resultados comuns acerca das hipóteses de interesses. Em experimento avaliado por Vieira et al. (2007), no qual avaliaram o uso da leucena como pastagem direta no campo, o resultado indicado pelo critério de Roy diferenciou-se dos demais ao rejeitar a hipótese nula de paralelismo entre dois fatores de interesse no experimento.

Na aplicação aqui realizada, em todos os testes utilizados, verificou-se a rejeição de igualdade do efeito dos tratamentos.

Observa-se que o proc GLM apresenta limitações a experimentos com presença de parcelas perdidas, que não é o caso do experimento em questão.

Ao usar o proc GLM é necessário que seja indicado um tipo de transformação e, com esse procedimento, tornou-se possível realizar a análise de regressão polinomial para os pesos dos animais, nas semanas avaliadas, em cada unidade experimental (animal). Com a análise de regressão, foi estudada a significância dos coeficientes dos polinômios no intuito de selecionar os parâmetros de efeito aleatório no modelo linear misto. Com o proc GLM, também foi possível confirmar o grau do polinômio sugerido pela análise exploratória, por meio da análise de regressão ajustada aos pesos médios dos animais ao longo das semanas avaliadas, para cada dieta alimentar. Com este resultado, foi identificado o modelo polinomial que representou a curva média ajustada pelo modelo linear misto para cada dieta alimentar nas semanas avaliadas.

O proc MIXED, ao contrário do proc GLM, permite a inclusão de indivíduos (parcelas experimentais) que tenham alguma observação perdida.

#### **4.4 Análise por meio do modelo linear misto**

A seguir, estão subdivididos, por tópicos, os passos fundamentais para proceder o ajuste das curvas polinomiais de crescimento por meio do modelo linear misto.

##### **4.4.1 Escolha dos efeitos fixos associados aos perfis médios de resposta da variável peso**

Sendo semanas, dietas e a interação dietas×semanas os componentes fixos do modelo linear, quando o objetivo da análise é ajustar curvas polinomiais de crescimento, o passo inicial é diagnosticar o grau do polinômio que representará a variação do peso médio do animal ao longo do tempo.

O grau do polinômio a ser ajustado pelo modelo linear misto para representar o comportamento do peso médio do animal ao longo das semanas avaliadas, para as seis dietas, foi o de segundo grau. As análises de regressões realizadas (Tabela 7) confirmaram o polinômio de segundo grau, que havia sido sugerido pelas Figuras 1 e Figura 5, pois os parâmetros de todos os polinômios ajustados foram significativos, a 1% de probabilidade, pelo teste *t* e o coeficiente de determinação superior a 97%, em todos os casos.

Tabela 7 Estimativas, erros padrões, valor do teste  $t$ , nível de significância (valor-p) para os parâmetros dos modelos das curvas polinomiais ajustadas para pesos dos animais (em gramas), no período de 15 semanas e coeficiente de determinação para as seis dietas.

NORM					
Coeficiente	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-265,66404	15,02	-17,68	<0,001	0,99
Linear	87,14728	2,99	29,08	<0,001	
Quadrático	-2,54770	0,13	-18,98	<0,001	
NCAF					
Coeficiente	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-225,87265	16,75	-13,48	<0,001	0,97
Linear	74,71879	3,34	22,36	<0,001	
Quadrático	-2,14489	0,14	-14,33	<0,001	
DREC					
Coeficiente	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-168,02485	24,34	-6,90	<0,001	0,99
Linear	47,38071	4,85	9,75	<0,001	
Quadrático	-0,85108	0,21	-3,91	<0,002	
DRECAF					
Coeficiente	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-162,06356	27,47	-5,9	<0,001	0,97
Linear	45,44891	5,48	8,28	<0,001	
Quadrático	-0,80572	0,24	-3,28	<0,004	
DESN					
Coeficientes	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-143,12730	17,97	-7,96	<0,001	0,98
Linear	48,08090	3,58	13,41	<0,001	
Quadrático	-1,32694	0,16	-8,26	<0,001	
DCAF					
Coeficientes	Estimativa	Erro padrão	Valor $t$	Valor - p	$R^2$
Intercepto	-128,14639	8,27	-15,50	<0,001	0,98
Linear	42,24351	1,64	25,50	<0,005	
Quadrático	-1,07790	0,07	-14,59	<0,001	

Outro passo muito importante no ajuste de curvas de crescimento é a identificação dos efeitos aleatórios, pois estes estarão associados à variação aleatória das curvas polinomiais de crescimento de cada unidade experimental.

#### **4.4.2 Identificação dos efeitos aleatórios associados aos perfis individuais de resposta da variável peso**

A identificação dos efeitos aleatórios, ou seja, os coeficientes (parâmetros) dos polinômios associados às unidades experimentais, pode ser realizada por meio gráfico, semelhante ao que foi feito no tópico anterior e/ou confirmando a escolha do grau do polinômio escolhido, analisando a significância de seus coeficientes pelo teste  $t$  por meio da análise de regressão em cada unidade experimental. Portanto, foram analisados, nesta etapa, os pesos em cada unidade experimental e não o peso médio, como feito anteriormente.

Pela impressão do teste  $t$  pelo proc GLM, pode ser confirmado que, apenas nas unidades experimentais 46 e 49, os coeficientes quadráticos não foram significativos, a 5% de probabilidade, pelo teste  $t$ . Pelo exposto, um polinômio de 2º grau representa adequadamente a relação funcional entre o peso do animal e o tempo em cada unidade experimental. Sendo assim, o intercepto e o coeficiente linear e quadrático das curvas polinomiais podem ser considerados de efeitos aleatórios.

A penúltima etapa antes da estimação dos parâmetros das curvas polinomiais de crescimento é a escolha das estruturas das matrizes de covariâncias mais adequadas, associada aos parâmetros de efeitos aleatórios e ao erro do modelo linear misto.

#### **4.4.3 Ajuste das matrizes de covariâncias, testes, estimação dos parâmetros e diagnóstico do modelo**

Foi considerado o modelo maximal, com todos os parâmetros, em todos os casos, assim como foi considerada, inicialmente, a estrutura VC (componente de variância) para  $R$ , que está associada ao erro do modelo 2.20, variando-se apenas a escolha da matriz de covariâncias para a variação entre indivíduo, ou seja, a escolha da estrutura de  $G$  do modelo 2.20. Após a escolha da estrutura de  $G$ , fixou-se a estrutura selecionada, variando-se apenas a estrutura de  $R$ . A estimação dos parâmetros das estruturas de covariâncias deu-se pelo método de máxima verossimilhança restrita (MVR) e observa-se que, em alguns casos, não houve convergência dos seus parâmetros. Todas as etapas anteriores foram implementadas no proc mixed do SAS, conforme rotinas apresentadas nos Anexos.

Conforme pode ser verificado na Tabela 8, tanto o critério de AIC quanto o de BIC, indicaram como mais adequada à estrutura UN (não estruturada) para  $G$  (associada aos parâmetros de efeitos aleatórios) e VC (componente da variância) para  $R$  (associada ao erro aleatório do modelo) por apresentarem os menores valores. Portanto, em função de ambos os testes indicarem as mesmas estruturas de covariâncias, não se faz necessário utilizar o teste assintótico da razão de verossimilhança.

Os critérios de AIC e BIC podem não ser concordantes ao serem favoráveis a modelos com estruturas de covariâncias distintas. Em experimento no qual se avaliou o peso de frangos, por exemplo, o critério de AIC favoreceu a opção da estrutura UN e o BIC à estrutura ARH(1) (BARBOSA, 2009).

Após ter selecionado a estrutura UN para a matriz de covariância associada aos efeitos aleatórios do modelo linear misto e VC para a matriz de covariância associada ao erro do modelo, procede-se à última etapa, que é a

estimação dos parâmetros das curvas polinomiais de crescimento do peso médio do animal, ao longo das 15 semanas de avaliação.

Tabela 8 Estimativas dos critérios de seleção das estruturas da matriz de covariâncias para G, variação entre indivíduos e R, variação intra indivíduo, considerando o modelo maximal e o método de estimação da máxima verossimilhança restrita (MVR).

Estruturas		Critérios		
R	G	$-2\log \text{ver.}$	AIC	BIC
VC	VC	7400,3	7400,3	7416,7
VC	CS	7546,9	7550,9	7555,1
VC	UN	7331,7	<b>7343,7</b>	<b>7356,3</b>
VC	AR	7546,9	7550,9	7555,1
VC	ARMA	7546,9	7550,9	7555,1
VC	TEOEP	7519,3	7525,3	7531,5
VC	CSH	7349,6	7359,6	7376,1
VC	HF	7352,0	7362,0	7372,5
CS	UN	7456,9	7460,9	7465,1
UN	UN		Não convergiu	
TEOEP	UN		Não convergiu	
CSH	UN		Não convergiu	
HF	UN		Não convergiu	

Utilizando-se a estrutura de covariâncias UN, sem estrutura definida, e VC, que se caracteriza por variâncias iguais e observações independentes, procede-se ao diagnóstico da normalidade dos resíduos gerados pelas curvas ajustadas para cada uma das seis dietas alimentares (Figura 6).

O diagnóstico do ajuste da modelagem das curvas polinomiais de crescimento dos pesos médios dos animais pode ser visualizado na Figura 6, a qual mostra uma tendência linear dos resíduos, não notando valores atípicos e também pela aplicação do teste de Shapiro-Wilks, em que foi obtido o valor  $W = 0,9902$  e o valor  $-p$  de 0,76, confirmado a normalidade dos resíduos.

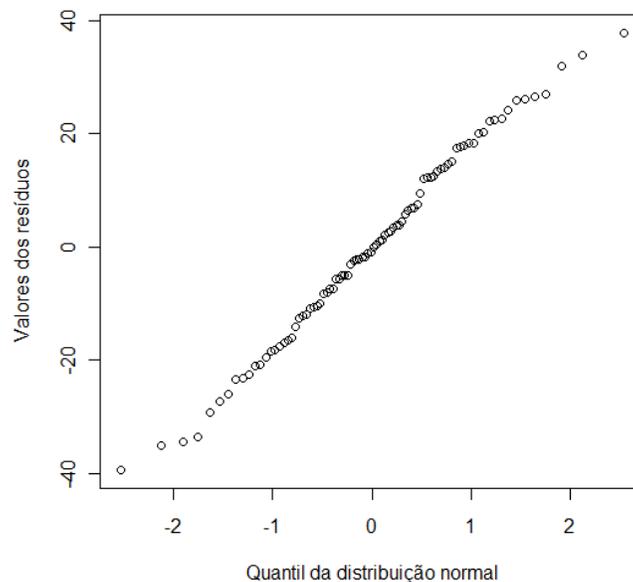


Figura 6 Representação gráfica da probabilidade normal dos resíduos das curvas ajustadas ao peso, em gramas, das dietas alimentares, na evolução das 15 semanas avaliadas

Após ter confirmado a normalidade dos resíduos das curvas estimadas, aplica-se o teste nos efeitos fixos do modelo, conforme mostrado na Tabela 9, em que todas as fontes de variação são significativas e, conseqüentemente, devem permanecer no modelo.

Tabela 9 Teste para os efeitos fixos para o ajuste do modelo selecionado dos pesos dos animais, em gramas, aferidos no período de 15 semanas, a partir da 4ª semana de vida, para as 6 dietas.

Fontes de variação	G.L. Num <sup>1</sup> .	G.L. Den <sup>2</sup> .	F	Valor - p
Dietas	5	54	60,61	<0,0001
Semanas	1	54	4241,92	<0,0001
Dietas *Semanas	5	54	74,60	<0,0001
Dietas*Semanas <sup>2</sup>	6	54	327,13	<0,0001

1: Graus do numerador. 2: Graus de liberdade do denominador

As comparações entre alguns contrastes de interesse prático (Tabela 10) sugerem que a inclusão do extrato aquoso de café foi significativa apenas na comparação dos tratamentos NORM *vs* NCAF. Apesar do fato de não ter sido diagnosticada diferença significativa entre os efeitos médios dos pesos dos animais presentes nas comparações das dietas DCAF *vs* DESN e DREC *vs* DRECAF, as dietas DCAF e DRECAF, que receberam a inclusão do extrato aquoso de café, têm menor valor médio, observando a grandeza numérica dos pesos em ambas as comparações.

Como na comparação das dietas NORM *vs* NCAF foi diagnosticada diferença significativa entre os valores médios das receptivas dietas e como a dieta que recebeu inclusão do extrato aquoso de café na dieta alimentar dos animais apresentou menor valor referente aos pesos médios dos animais, é

razoável afirmar que, neste caso, a inclusão do café influenciou a diminuição dos pesos dos animais.

O uso dos modelos lineares mistos confirmou os resultados obtidos por Barbosa (2007), ou seja, que o uso do café na dieta alimentar dos ratos que receberam a dieta NCAF diminuiu o peso corporal desses animais, comparado ao grupo de animais que receberam a dieta NORM.

Tabela 10 Estimativas e avaliação de contrastes formados por dietas que receberam e não receberam extrato aquoso de café, utilizando a estatística *t*.

Contraste	Estimativa	Erro padrão	Graus de liberdade	Valor <i>t</i>	Valor - p
DCAF <i>vs</i> DESN	-14,4480	10,38	54	- 1,39	0,1698
DREC <i>vs</i> DRECAF	8,9533	10,38	54	0,86	0,39
NORM <i>vs</i> NCAF	40,6620	10,38	54	3,92	0,0005
Médias: DCAF = 185,99; DESN=200,43; DREC =234,30; DRECAF 225,34; NORM = 337,13; NCAF = 296,46					

Na Tabela 11 apresenta-se o teste referente aos parâmetros das curvas médias de crescimento dos pesos dos grupos de animais ao longo das 15 semanas. Observa-se que a curva a ser ajustada para a dieta NORM constará da menor estimativa de intercepto e coeficiente quadrático entre todas as dietas.

Tabela 11 Estimativas e teste para os parâmetros do modelo das curvas polinomiais ajustadas pelo modelo linear misto para pesos dos animais, em gramas, aferidos no período de 15 semanas, a partir da 4ª semana de vida, para as 6 dietas alimentares.

Efeitos	Dietas	Estimativas	Erro padrão	Graus de liberdade	Valor <i>t</i>	Valor - p
Intercepto		-225,87	6,7797	54	-33,32	<0,0001
Dietas	DCAF	97,7263	9,5879	54	10,19	<0,0001
Dietas	DESN	82,7454	9,5879	54	8,63	<0,0001
Dietas	DREC	57,8478	9,5879	54	6,03	<0,0001
Dietas	DRECAF	63,8091	9,5879	54	6,66	<0,0001
Dietas	NORM	-39,7914	9,5879	54	-4,15	<0,0001
Dietas	NCAF	0				
Semanas		74,7188	2,1627	54	34,55	<0,0001
Sem * DCAF	DCAF	-32,4753	3,0585	54	-10,62	<0,0001
Sem * DESN	DESN	-26,6379	3,0585	54	-8,71	<0,0001
Sem * DREC	DREC	-27,3381	3,0585	54	-8,94	<0,0001
Sem * DRECAF	DRECAF	-29,2699	3,0585	54	-9,57	<0,0001
Sem * NORM	NORM	12,4285	3,0585	54	4,06	<0,0002
Sem * NCAF	NCAF	0				
Sem <sup>2</sup> * DCAF	DCAF	-1,0779	0,08854	54	-12,17	<0,0001
Sem <sup>2</sup> * DESN	DESN	-1,3269	0,08854	54	-14,99	<0,0001
Sem <sup>2</sup> * DREC	DREC	-0,8511	0,08854	54	-9,61	<0,0001
Sem <sup>2</sup> * DRECAF	DRECAF	-0,8057	0,08854	54	-9,10	<0,0001
Sem <sup>2</sup> * NORM	NORM	-2,5477	0,08854	54	-28,77	<0,0001
Sem <sup>2</sup> * NCAF	NCAF	-2,1449	0,08854	54	-24,22	<0,0001

Na Figura 7 encontram-se a representação gráfica e as estimativas das equações polinomiais de segundo grau para as seis dietas. Observa-se que a dieta NORM tem o maior peso médio dos grupos de animais e a dieta DCAF, o menor peso médio, a partir da 7ª semana, aproximadamente. Também é perceptível que se comparando as dietas NCAF, DRECAF e DCAF, que receberam a inclusão do extrato aquoso de café na dieta alimentar, dois a dois, na ordem respectiva, com as dietas de animais NORM, DREC e DESN, vê-se que as dietas que receberam o café tenderam a diminuir o ganho médio de peso ao longo das 15 semanas e esse comportamento tendeu a aumentar ao longo do tempo. O ganho de peso foi mais evidente na dieta NORM sobre a dieta NCAF e não significativo na dieta DESN sobre DCAF, assim como na dieta DREC sobre DRECAF. As curvas polinomiais ajustadas para cada dieta alimentar foram as seguintes:

$$\text{NORM, } \hat{y} = -265,6614 + 87,1473 t - 2,5477 t^2 ,$$

$$\text{NCAF, } \hat{y} = -225,87 + 74,7188 t - 2,1449 t^2 ,$$

$$\text{DREC, } \hat{y} = -168,0222 + 47,3807 t - 0,8511 t^2 ,$$

$$\text{DRECAF, } \hat{y} = -162,0609 + 45,4549 t - 0,8057 t^2$$

$$\text{DESN, } \hat{y} = -143,1246 + 48,0809 t - 1,3269 t^2 \text{ e}$$

$$\text{DCAF, } \hat{y} = -128,1437 + 42,2435 t - 1,0779 t^2 .$$

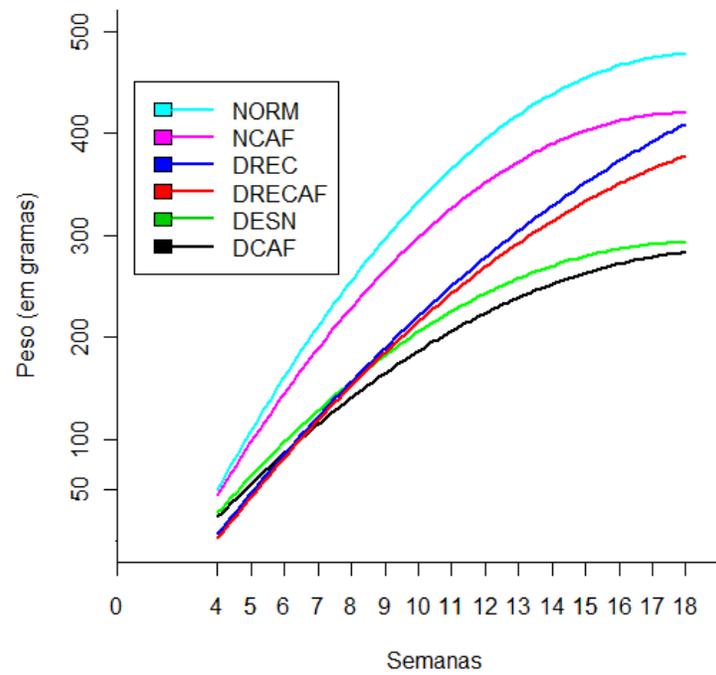


Figura 7 Descrição gráfica do crescimento de ratos por meio de curvas polinomiais ajustadas ao peso por animal (em gramas), durante período de 15 semanas, para cada dieta alimentar

## 5 CONCLUSÕES

A metodologia de modelo linear misto mostra-se como a melhor metodologia, dentre as estudadas, por apresentar versatilidade no ajuste de curvas polinomiais de crescimento, ao tornar possível explorar e captar, de forma eficiente, informações presentes nos dados, por meio de seleção da estrutura da matriz de covariâncias.

As análises estatísticas realizadas produziram os mesmos resultados aos parâmetros de efeitos fixos, pelos testes aplicados nas diferentes metodologias.

A inclusão de café nas dietas alimentares tende a diminuir o peso dos animais (ratos) somente associado a dieta (NORM), que recebeu ração de biotério enriquecida com ração para cães.

## **6 TRABALHOS FUTUROS**

Este trabalho não se encerra com os resultados já obtidos. No intuito de aperfeiçoá-lo, pretende-se ainda:

- 1) estudar a metodologia de modelos mistos para casos em que se faz necessário relacionar a variável resposta de algum fator analisado com o tempo, porém, com a utilização de curvas não-lineares;
- 2) estudar, por meio de simulação, a aplicabilidade das técnicas uni e multivariadas de perfil, bem como o teste de Mauchly, sobretudo em casos de desbalanceamentos.

## REFERÊNCIAS

AMADOR, J. P. **Modelos mistos no ajuste de curvas de crescimento de *Ricinus communis* L.** 2010. 139 p. Tese (Doutorado em Agronomia) - Universidade Federal de Santa Maria, Santa Maria, 2010.

ANDRADE, D. F.; SINGER, J. M. Análise de dados longitudinais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7., 1986., Campinas. **Anais...** Campinas: SINAPE, 1986. p. 106.

APARICIO, P. D. et al. Controle da matocompetição em plantios de dois clones de *Eucalyptus* × *urograndis* no amapá. **Ciência Florestal**, Santa Maria, v. 20, n. 3, p. 381-390, 2010.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE CAFÉ. **Indicadores da indústria de café no Brasil.** Disponível em: <<http://www.abic.com.br>>. Acesso em: 10 nov. 2011.

BARBOSA, M. **Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos.** 2009. 118 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2009.

BARBOSA, W. M. **Efeito da ingestão de café sobre a liberação de glutamato sinaptossomal, estimulada por toxina escorpionic, em ratos.** 2007. 60 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

BOECK, P.; NABEREZNY, C. L.; TAVARES, H. R. **Linear nonlinear generalized mixed models: inference and applications.** Fortaleza: Escola de Modelos de Regressão, 2011. 123 p.

BOX, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems II: effects of inequality of variance and of correlation between errors in the two-way classification. **Annals of the Mathematical Statistics**, Ann Arbor, v. 25, p. 484-498, 1954.

CECON, P. R. et al. Análise de medidas repetidas na avaliação de clones de café 'Conilon'. **Pesquisa Agropecuária Brasileira**, Brasília, v. 43, n. 9, p. 1171-1176, set. 2008.

ENDE, C. N. von. Repeated-measures analysis: growth and other time-dependent measures. In: SCHEINER, S. M.; GUREVITCH, J. (Ed.). **Design and analysis of ecological experiments**. New York: Chapman and Hall, 1993. p. 113-137.

FARAWAY, J. J. **Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models**. New York: Chapman & Hall/CRC, 2006. 331 p.

FERREIRA, D. F. **Análise multivariada**. Lavras: UFLA, 2008. 662 p.

FREITAS, E. G. **Análise de dados longitudinais em experimentos com cana-de-açúcar**. 2007. 70 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2007.

FREITAS, E. G. et al. Modelo univariado aplicado a dados longitudinais. **Revista Brasileira Biometria**, São Paulo, v. 26, n. 2, p. 93-106, 2008.

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

GOMES, G. B. **Efeito do café sobre o sistema de biotransformação hepático de ratos durante o crescimento e desenvolvimento**. 2004. 96 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

HAIR, J. J. F. et al. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.

HAUSER, L. et al. Curva de crescimento usando modelo misto: uma aplicação na progressão da doença de Machado-Joseph. **Revista do Hospital de Clínicas de Porto Alegre**, Porto Alegre, v. 29, n. 1, p. 5-17, 2009.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, Bethesda, v. 31, p. 423-449, 1975.

\_\_\_\_\_. Estimation of changes in herd environment. **Journal of Dairy Science**, Champaign, v. 32, p. 706-711, 1949.

\_\_\_\_\_. Selection index and expected genetic advance. In: HANSON, W. D.; ROBINSON, A. F. (Ed.). **Statistical genetics and plant breeding**. Washington: NAS-NRC, 1963. p. 141-163.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Bethesda, v. 15, p. 192-193, 1959.

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, 1970.

\_\_\_\_\_. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. **Journal of Educational Statistics**, Alexandria, v. 1, n. 1, p. 69-82, 1976.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New York: Prentice Hall, 2007. 773 p.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, Washington, v. 38, p. 963-997, 1982.

LITTELL, R. C. et al. **SAS mixed models**. 2<sup>nd</sup> ed. Cary: SAS Institute, 2006. 814 p.

LOUGHIN, T. M. et al. On the analysis of long-term experiments. **Journal of the Royal Statistical Society: Series A**, London, v. 170, n. 1, p. 29-42, 2007.

MACHADO, L. M. M.; DÓREA, J. G.; COSTA, T. H. M. **Aspectos econômicos, nutricionais e funcionais do café**. Disponível em: <[http://www.nutricaoempauta.com.br/lista\\_artigo.php?cod=497](http://www.nutricaoempauta.com.br/lista_artigo.php?cod=497)>. Acesso em: 10 nov. 2011.

MALHEIROS, E. B. Precisão da análise de experimentos com medidas repetidas no tempo usando procedimentos do SAS. **Revista de Matemática e Estatística**, São Paulo, v. 19, n. 2, p. 253-272, 2001.

\_\_\_\_\_. Precisão de teste F univariados usados em experimentos com medidas repetidas no tempo, quando a condição de esfericidade da matriz de covariâncias não é verificada. **Revista de Matemática e Estatística**, São Paulo, v. 22, n. 2, p. 23-29, 2004.

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

MCCULLOCH, C. E.; SEARLE, S. R. **Generalized, linear, and mixed models**. 2<sup>nd</sup> ed. New York: J. Wiley, 2001. 325 p.

MEREDITH, M. P.; STEHMAN, S. V. Repeated measures experiments in forestry: focus on analysis of response curves. **Canadian Journal of Forest Research**, Ottawa, v. 21, n. 7, p. 957-965, July 1991.

MONTGOMERY, D. C. **Design and analysis of experiments**. New York: J. Wiley, 2001. 684 p.

MOTULSKY, H.; CHRISTOPOULOS, A. **Fitting models to biological data using linear and nonlinear regression**. San Diego: Graph Pad, 2003. 351 p.

NOBRE, J. S.; SINGER, J. M. Residuals analysis for linear mixed models. **Biometrical Journal**, Berlin, v. 49, n. 6, p. 863-875, 2007.

PERRI, S. H. V.; IEMMA, A. F. Procedimento "MIXED" do SAS® para análise de modelos mistos. **Scientia Agricola**, Piracicaba, v. 56, n. 4, p. 959-967, 1999.

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

POTTHOFF, R. F.; ROY, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrics**, Bethesda, v. 31, p. 313-326, 1964.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2011. Disponível em: <<http://www.R-project.org>>. Acesso em: 15 nov. 2011.

ROCHA, E. B. **Modelos não-lineares para dados longitudinais provenientes de experimentos em blocos casualizados abordagem bayesiana**. 2010. 96 p. Dissertação (Mestrado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2010.

RYAN, L.; HATFIELD, C.; HOFSTTER, M. Caffeine reduces time-of-day effects on memory performance in older adults. **American Psychological Association**, Washington, v. 13, n. 1, p. 68-71, 2002.

SEARLE, S. R. **Linear models**. New York: J. Wiley, 1971. 532 p.  
SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York: J. Wiley, 1992. 501 p.

SILVA, S. A. A. **Análise de experimento com medidas repetidas**: uma aplicação no efeito da ingestão do café no controle de peso. 2009. 51 f. Dissertação (Mestrado em Estatística) - Universidade Vale do Rio Verde de Três Corações, Três Corações, 2009.

SINGER, J. M.; ROCHA, F. M. M.; NOBRE, J. S. Análise de medidas repetidas. In: JORNADA DE ESTATÍSTICA, 4., 2004, Maringá. **Anais...** Maringá: UEM, 2004. p. 120.

SMITH, A. et al. Caffeine and central Noradrenaline: effects on mood, cognitive performance, eye movements and cardiovascular function. **Journal of Psychopharmacology**, Lausanne, v. 17, n. 3, p. 283-292, Mar. 2003.

STORCK, L. et al. **Experimentação vegetal**. 2. ed. Santa Maria: UFSM, 2006. 198 p.

SYSTEM ANALYSIS STATISTICAL INSTITUTE. **SAS onlinedoc® 9.1.3**. Cary, 2004. Disponível em:  
<<http://support.sas.com/onlinedoc/913/docMainpage.jsp>>. Acesso em: 10 dez. 2010.

VIEIRA, F. T. P. A. et al. Uma abordagem multivariada em experimentos Silvopastoral com *Leucaena leucocephala* (Lam.) de Wit no Agreste de Pernambuco. **Ciência Florestal**, Santa Maria, v. 17, n. 4, p. 333-342, 2007.

XAVIER, L. H.; DIAS, C. T. S. Acurácia do modelo univariado para análise de medidas repetidas por simulação multidimensional. **Scientia Agricola**, Piracicaba, v. 58, n. 2, p. 241-250, 2001.

## ANEXOS

### ANEXO A

Rotinas utilizadas nas análises dos modelos univariado, multivariado e linear misto dos programas R e SAS®.

```
##### Programas R #####

rm(list=ls(all=TRUE))

dados<-read.table("Dados5.txt", header=TRUE)

attach(dados)

# Bibliotecas necessárias
require(lattice);require(psych);require(car) ;require(nlme)

grafico1<-groupedData(Peso~Tempo|Tratamento,data=dados,order. groups=F)
plot(grafico1,xlab="Semanas",ylab="Peso(gramas)")
boxplot(Peso~Parcela,xlab="Unidades      experimenatais",ylab="Peso      (em
gramas)")
boxplot(Peso~Tempo:Tempo,col = "yellow",xlab="Semanas",ylab="Peso (em
gramas)")
axis(1, at=8:9, lab=c("11","12")) ;axis(1, at=12:13, lab=c("15","15")); axis(1,
at=10:11, lab=c("13","13")); axis(1, at=14:15, lab=c("17","18"))
boxplot(Peso~Tratamento,subset  = Tratamento  ==  "NORM",col  =
"yellow",xlab="Grupos",ylab="Peso (em gramas)")
boxplot(Peso~Tratamento,subset  = Tratamento  ==  "NCAF",col  =
"brown",add=TRUE)
boxplot(Peso~Tratamento,subset  = Tratamento  ==  "DREC",col  =
"blue",add=TRUE)
```

```

boxplot(Peso~Tratamento,subset = Tratamento == "DRECAF",col =
"brown",add=TRUE)
boxplot(Peso~Tratamento,subset = Tratamento == "DESN",col =
"red",add=TRUE)
boxplot(Peso~Tratamento,subset = Tratamento == "DCAF",col =
"brown",add=TRUE)
axis(1, at=4:5, lab=c("DRECAF","NCAF"))
axis(1, at=4:5, lab=c("DRECAF","NCAF"))
interaction.plot(Tempo,Tratamento,Peso,ylab="Peso
médio(gramas)",xlab="Semanas",col= 2:5,bty="l",xaxs="i")
axis(1,at=1:15, lab=c("4","5","6","7","8","9","10","11","12","13","14",
"15","16","17","18"))
axis(1, at=7:8, lab=c("10","10"))
axis(1, at=9:10, lab=c("12","12"))
axis(1, at=11:12, lab=c("14","14"))
axis(1, at=13:14, lab=c("16","16"))
axis(1, at=15, lab=c("18"))
axis(2, at=50:51, lab=c("50","51"))

##### Análise Univariada #####
#####
anova<-aov(Peso~Tratamento*factor(Tempo)+Error(factor(Parcela)))
summary(anova)
##### Função para calcular o teste de Mauchly e as estimativas dos ajustes dos
graus de liberdade
esfericidade1_wederson = function(Sigma,g,N,alpha)
{
#g <- número de níveis do fator em estudo

```

```

#N <- número de parcelas
#aipha = nível de significância
  if(is.null(Sigma))
    stop("Sigma deve ser indicado")
  n <- ncol(Sigma)
  T<-ncol(Sigma)
  W<-(((T)^(T))*det(Sigma))/(tr(Sigma)^(T))
  p=ncol(Sigma)
  n1<-(N/g-1)*g
  f<-(p*(p+1)/2)-1
  d<-1-(2*p^2+p+2)/(6*p*n1)
  lambda<--n1*d*log(W)
  p_valor<-(1-pchisq(-n1*d*log(W),f))
  if(p_valor >= alpha)a<-"Matrix de CV atende H-F, aceito H0" else
  a="Matrix de CV 'NÃO' atende H-F, rejeito H0"
  eig <- eigen(Sigma)$values
  GG <- (sum(eig)^2)/((n)*sum(eig^2))
  HF <- (N*(n)*GG - 2) / ((n)*(N - g - ((n)*GG)))
  B <- 1/(n)
  epsilon <- c(GG, HF, B)
  names(epsilon) <- c("Geisser-Greenhouse épsilon", "Huynh-Feldt
  épsilon", "Box épsilon")
  result
  list(Graus_de_liberdade_f=f,Teste_Mauchly_W=W,Teste_Mauchly_estatística=
  lambda,Teste_Mauchly_p_valor=p_valor,Teste_Mauchly_Classificação=a,
  Épsilon = epsilon)
  class(result) <- "esfer"
  return(result)

```

```

}
#####
##### Função que imprime as fontes de variação referente a subparcela da
ANOVA com o teste F com os graus de liberdade corrigido.
ANOVA_Ajustada_wederson <- function(x, y)
{
  if(!inherits(x, "aovlist"))
    stop("x deve ser da classe 'aovlist'")
  if(!inherits(y, "esfer"))
    stop("x deve ser da classe 'esfer'")
  R <- summary(x)$"Error: Within"[[1]]
  n <- nrow(R)
  eGG <- y$Épsilon[1]
  eHF <- if(y$Épsilon[2] > 1) 1 else y$Épsilon[2]
  eB <- y$Épsilon[3]
  for(i in 1:(n-1)){
    R$"Pr(>F)_GG"[i] <- 1 - pf(R$"F value"[i], R$Df[i]*eGG,
    R$Df[n]*eGG)
    R$"Pr(>F)_HF"[i] <- 1 - pf(R$"F value"[i], R$Df[i]*eHF,
    R$Df[n]*eHF)
    R$"Pr(>F)_B"[i] <- 1 - pf(R$"F value"[i], R$Df[i]*eB, R$Df[n]*eB)}
  R$"Pr(>F)_GG"[n] = R$"Pr(>F)_HF"[n] = R$"Pr(>F)_B"[n] = NA
  return(R)
}
#####
##### Função que verifica matematicamente a condição H-F.
H_F<-function(x)
{

```

```

y<-matrix(c(0),ncol(x),ncol(x))
lambda=abs(tr(x)/(ncol(x))-(sum(x)-tr(x))/((ncol(x)^2)-length(diag(x))))
  for(i in 1:ncol(x)-1)
  {
    for(j in i:ncol(x)-1)
    {
      y[i,j]=x[i,i]+x[(j=j)+1,(j=j)+1]-2*x[i,(j=j)+1]
    }
  }
y<-c(y)
y<-(y)[y==2*lambda]
  if(length(y)==choose(ncol(x),2))a="Matrix de CV atende a condição H-
F" else a="Matrix de CV NÃO atende a condição H-F"
  return(list(Valores_variâncias=y,Lambda=lambda,Classificação=a))
}
Sigma<-matrix(c(a,b),14,14)# impressa no SAS no comando multivariado
#####
##### Rodando as funções construídas
H_F(Sigma)
wederson<-esfericidade1_wederson(Sigma,6,60,0.05);wederson
ANAVA_Ajustada_wederson (anova, wederson)
#####
##### Programa SAS® #####
dm 'output; clear; log; clear;';
options formdlim='*' pageno=001 ls=100;
data Wederson (keep=UE Grupo Rep S4 S5 S6 S7 S8 S9 S10
S11 S12 S13 S14 S15 S16 S17 S18 )
Leandro(keep=UE Grupo Rep Semanas W );

```

```

input UE Grupo  Rep S4 S5 S6 S7 S8 S9 S10 S11 S12 S13
S14 S15 S16 S17 S18  ;
output Wederson;
W=S4; Semanas=4; output Leandro;
W=S5; Semanas=5; output Leandro;
W=S6; semanas=6; output Leandro;
W=S7; Semanas=7; output Leandro;
W=S8; Semanas=8; output Leandro;
W=S9; Semanas=9; output Leandro;
W=S10; Semanas=10; output Leandro;
W=S11; Semanas=11; output Leandro;
W=S12; Semanas=12; output Leandro;
W=S13; Semanas=13; output Leandro;
W=S14; Semanas=14; output Leandro;
W=S15; Semanas=15; output Leandro;
W=S16; Semanas=16; output Leandro;
W=S17; Semanas=17; output Leandro;
W=S18; Semanas=18; output Leandro;
; cards;

```

1	1	1	28.6	51.6	79.6	117.9	(...)	278	287.9
2	1	2	30.2	53.7	85.2	104.7	(...)	306.3	316.5
3	1	3	28.9	52.8	83.5	103.7	(...)	268.4	278.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
59	6	9	31.2	57.2	94	129.5	(...)	390.1	397.8
60	6	10	32.4	57.5	91.7	133.3	(...)	392.6	395.3

```

;
proc mixed data=Leandro;
class Grupo Rep ;
/*Efetua a análise Multivariada*/
proc glm data=Wederson;
class Grupo;
model S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16
S17 S18 = Grupo /nouni;
repeated Semanas 5 polynomial/summary printm printe;
run; quit;

/*Cria um polinômio de 2° grau para cada tratamento e
aplica o teste t, e também imprime o coeficiente de
determinação*/
proc sort; by Grupo Semanas; proc means; by Grupo
Semanas; Var W;
output out = polim mean=W;
proc print data=polim;
data polim1 ; set polim;
Semanas2=Semanas*Semanas;
proc glm;
model W=Semanas Semanas2; by Grupo;
run; quit;

/*Faz o ajuste da matriz de covariância, apenas deve-
se substituir no comando type em randon o comando das
estruturas. Também estima os parâmetros das curvas de
crescimento e aplica os testes */
proc mixed data=Leandro covtest method=REML;
class Grupo Rep;

```

```
model W= Grupo semanas Grupo*Semanas
Grupo*Semanas*Semanas /chisq s;
repeated / type=vc subject=Rep(Grupo) r;
random intercept Semanas Semanas*Semanas/ type= UN
subject=Rep(Grupo);
run;
quit;

/* Para realizar o teste t para contrastes de
interesse, acrescenta-se ao comando anterior: */
lsmeans Grupo/ pdiff adjust=tukey;

/* Gera a análise de regressão para cada uma das 60
unidades experimentais, com o objetivo de diagnosticar
os parâmetros de efeitos aleatórios: */

data Ferreira ; set Leandro;
proc glm data =Ferreira;
model W= Semanas Semanas*Semanas/ss3; by UE;
```

## ANEXO B

Tabela 1A Dados reais referentes aos pesos dos ratos, em gramas, da 4ª a 18ª semana de vida, dos grupos mononutridos, desnutridos e desnutridos, recuperados de um experimento desenvolvido na Universidade Federal de Minas Gerais (UFMG) (SILVA, 2009).

Grupo	Semanas														
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
DCAF	28,6	51,6	79,6	117,9	139,7	166,5	187,3	202,3	221,0	241,1	248,6	259,2	270,5	278,0	287,9
DCAF	30,2	53,7	85,2	104,7	151,0	193,9	228,6	221,0	243,6	256,5	265,0	270,1	297,0	306,3	316,5
DCAF	28,9	52,8	83,5	103,7	143,6	181,9	219,0	204,8	221,7	232,3	237,0	239,5	262,9	268,4	278,8
DCAF	26,8	51,5	75,3	112,5	131,8	155,9	175,0	189,8	204,8	223,2	232,3	238,0	245,8	251,2	260,0
DCAF	27,1	51,9	82,5	100,7	144,2	178,2	216,5	202,5	223,0	234,9	245,0	248,8	266,1	273,1	287,6
DCAF	27,0	53,2	75,8	112,6	132,9	157,5	177,4	194,5	215,2	229,3	240,6	252,0	260,2	263,0	275,0
DCAF	25,2	38,8	96,2	139,9	181,2	180,8	209,6	223,2	242,0	247,8	263,0	270,8	285,9	282,9	292,4
DCAF	26,7	53,2	75,1	113,8	135,2	161,7	176,4	194,0	216,2	233,1	246,3	260,2	271,6	280,0	292,4
DCAF	26,6	52,0	72,7	107,9	130,0	156,2	173,0	188,7	209,8	230,3	239,7	251,7	260,7	261,1	276,8
DCAF	28,8	57,4	82,1	119,1	142,9	169,9	192,2	208,5	232,3	256,7	268,5	285,0	298,5	304,3	314,7
DESN	39,4	71,4	104,0	122,4	168,1	200,8	247,1	233,1	256,3	280,1	287,6	292,0	302,0	312,2	318,2
DESN	38,3	74,5	105,9	123,3	179,4	211,2	259,0	244,1	256,3	261,8	269,9	279,3	296,6	304,0	313,2

TABELA 1A, Continuação

Grupo	Semanas														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DESN	37,9	72,4	105,2	124,6	173,6	201,2	242,1	233,1	254,9	257,2	272,8	281,0	296,1	305,1	312,9
DESN	36,0	71,6	105,9	122,6	173,2	205,5	251,2	237,3	260,3	271,8	286,0	290,0	306,0	309,3	321,7
DESN	33,1	55,7	82,4	103,3	153,8	181,5	229,0	213,2	228,1	236,3	247,4	260,0	270,0	280,0	279,0
DESN	32,3	53,3	80,3	104,5	147,9	176,1	216,2	210,6	228,3	234,6	250,4	253,5	269,2	277,2	278,8
DESN	31,7	48,5	80,9	102,4	151,5	180,4	215,9	212,6	231,0	241,2	257,8	264,0	276,0	286,9	290,4
DESN	35,5	57,2	86,3	107,8	155,4	183,8	230,5	219,2	236,1	247,8	258,3	265,5	280,5	292,0	285,3
DESN	33,5	55,2	80,3	102,1	150,4	178,3	220,5	212,4	228,9	238,9	249,4	257,0	268,0	277,6	283,8
DESN	31,8	46,8	83,4	109,1	144,3	172,3	198,8	213,6	230,0	244,8	265,6	274,0	281,7	288,3	295,8
DREC	30,0	43,4	84,6	106,9	142,4	170,0	223,7	261,8	299,0	321,5	346,1	361,5	378,3	387,3	395,0
DREC	32,9	43,9	80,0	107,0	143,5	176,7	218,9	249,0	281,5	316,6	340,9	353,9	359,3	377,0	380,0
DREC	32,0	44,2	79,2	105,5	137,6	163,0	203,8	247,5	276,3	315,9	336,3	355,1	366,9	372,3	382,1
DREC	30,5	45,6	82,8	104,2	138,2	165,4	206,6	238,0	268,5	322,3	316,3	331,0	345,2	357,6	371,2
DREC	30,6	44,2	83,1	107,1	146,0	170,8	216,6	253,9	286,3	320,4	351,4	371,9	400,0	410,6	424,8
DREC	29,1	43,8	78,5	104,6	139,5	167,0	210,0	235,8	264,3	283,4	304,2	314,6	335,0	341,6	344,3
DREC	29,0	43,9	81,7	104,2	140,4	174,0	223,0	266,7	304,6	347,6	371,7	387,6	399,3	416,5	415,5

TABELA 1A, Continuação

Grupo	Semanas														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DREC	33,8	50,7	89,1	109,5	148,4	179,0	232,4	269,3	303,5	339,6	370,2	375,5	393,5	406,0	413,2
DREC	31,7	49,1	89,7	109,8	143,7	171,5	225,8	271,5	299,7	322,1	355,0	367,6	393,1	407,6	423,2
DREC	33,4	48,0	83,7	101,4	137,0	165,7	208,7	251,6	292,6	315,3	347,6	360,1	377,0	384,5	390,1
DREC	29,8	50,2	80,0	101,8	127,6	152,0	196,2	236,5	272,2	291,5	322,6	336,7	345,4	365,0	368,2
AF DREC	29,3	50,1	85,6	106,6	136,0	164,0	225,3	272,0	312,5	349,5	372,5	394,5	411,6	424,2	441,9
AF DREC	30,4	47,8	84,0	103,6	130,8	156,7	212,3	253,5	290,0	319,6	338,9	357,6	377,0	381,0	398,0
AF DREC	27,4	43,2	73,7	102,6	129,4	155,7	219,9	253,6	285,3	304,2	330,0	342,1	355,2	367,0	373,8
AF DREC	30,0	49,1	83,8	104,7	132,8	159,1	207,5	260,0	262,2	326,8	353,2	359,1	351,3	377,0	390,5
AF DREC	30,0	45,6	78,5	102,4	128,0	155,2	197,2	234,0	304,1	283,1	307,7	318,5	337,1	347,5	362,2
AF DREC	26,2	44,1	76,6	97,1	126,8	151,8	206,6	246,0	275,6	305,5	332,8	341,4	357,0	363,5	368,6
AF DREC	33,0	44,8	79,5	103,6	132,0	158,5	221,1	282,0	320,6	350,3	396,3	411,2	435,7	444,5	450,6
AF DREC	30,2	43,6	75,6	97,9	126,3	148,1	188,2	202,0	218,4	225,3	230,2	242,5	252,9	255,7	255,4
AF DREC	31,0	45,7	80,0	101,4	131,1	150,8	207,2	247,5	290,3	315,0	345,0	360,6	367,8	386,7	391,8
AF NORM	50,4	107,6	170,4	199,5	284,8	347,0	372,0	400,3	430,9	451,8	469,0	480,0	488,4	488,5	480,1
NORM	45,0	100,0	161,6	194,4	275,4	343,0	370,1	395,7	426,2	445,1	472,0	482,0	483,5	498,5	492,4

TABELA 1A, Continuação

Grupo	Semanas														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NORM	58,2	112,9	172,6	199,8	282,6	337,0	368,1	393,1	415,9	432,7	447,2	450,5	472,2	470,3	456,5
NORM	62,0	124,5	188,3	223,8	300,5	357,0	397,5	415,0	452,5	463,3	490,1	496,0	504,0	500,3	498,2
NORM	51,7	113,4	171,8	207,6	295,0	359,0	384,2	424,6	452,3	477,8	504,0	512,0	534,4	553,1	545,0
NORM	59,8	111,3	167,1	188,6	266,1	320,9	346,2	370,3	404,8	421,2	442,8	452,0	460,9	470,8	463,5
NORM	59,8	113,1	176,0	208,2	290,4	346,2	373,7	410,1	432,0	442,0	460,1	478,0	483,0	486,8	494,0
NORM	43,6	86,4	123,6	170,9	228,0	274,5	314,5	346,6	379,1	406,8	423,5	431,3	443,4	453,7	466,9
NORM	44,2	90,0	139,2	184,0	238,0	281,2	318,2	343,0	381,8	404,8	420,3	431,7	440,1	414,6	445,5
NORM	48,1	92,8	141,3	186,5	240,2	86,7	324,8	354,0	380,3	408,1	429,0	436,7	442,5	443,9	451,0
NCAF	47,8	100,7	155,3	182,2	251,0	286,0	318,0	343,4	363,0	373,2	388,1	393,1	404,8	419,9	420,2
NCAF	52,9	94,6	147,2	170,4	236,5	279,7	313,0	336,6	354,0	364,0	377,2	379,0	390,1	395,6	390,1
NCAF	55,1	108,2	168,2	197,4	270,5	328,4	361,0	390,7	400,9	425,6	442,6	455,0	475,3	477,1	468,5
NCAF	46,1	94,1	148,5	175,3	245,6	287,0	316,5	344,0	364,2	370,5	379,4	378,3	400,2	399,6	396,3
NCAF	43,3	86,7	144,2	175,0	249,1	294,6	328,4	356,9	382,6	351,3	403,5	410,5	427,2	442,9	430,5
NCAF	48,3	90,1	148,2	177,4	258,3	310,0	345,5	370,7	404,1	413,8	433,0	441,7	456,5	477,6	482,0

TABELA 1A, Continuação

Grupo	Semanas														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NCAF	48,0	102,2	155,8	176,6	250,1	298,9	332,8	349,6	372,0	382,1	395,0	402,5	414,2	425,0	430,0
NCAF	58,0	111,8	166,2	193,6	268,9	309,1	343,0	367,6	396,0	410,2	430,0	443,0	457,5	469,1	481,3
NCAF	31,2	57,2	94,0	129,5	179,9	225,4	266,0	287,0	311,5	329,6	349,6	361,1	371,3	390,1	397,8
NCAF	32,4	57,5	91,7	133,3	180,4	220,0	260,0	279,3	309,2	329,4	345,3	360,3	371,9	392,6	395,3