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

Análise de dados longitudinais em experimentos com cana-de-açúcar

Edjane Gonçalves de Freitas

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

Piracicaba

2007

Edjane Gonçalves de Freitas
Engenheira Agrônoma

Análise de dados longitudinais em experimentos com cana-de-açúcar

Orientador:

Prof. Dr. **DECIO BARBIN**

Dissertação apresentada para obtenção do
título de Mestre em Agronomia. Área de
concentração: Estatística e Experimentação
Agrônômica

Piracicaba

2007

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Freitas, Edjane Gonçalves de
Análise de dados longitudinais em experimentos com cana-de-açúcar / Edjane
Gonçalves de Freitas. - - Piracicaba, 2007.
74 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Análise de dados longitudinais 2. Análise multivariada 3. Cana-de-açúcar 4. Medidas
repetidas I. Título

CDD 519.5

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

Dedicatória

A

Deus

Por ter me concedido essa graça.

A Nossa Senhora das Graças,
pela proteção, consolo e paz.

Aos meus pais, Antônio Gonçalves
e Pureza Maria, grandes guerreiros,
meus heróis, exemplos de força e luta
na batalha da vida.

AGRADECIMENTOS

Ao Prof^o Decio Barbin pelos ensinamentos, orientação, força e amizade, e pela pessoa humana sempre presente e disposto a ajudar.

Aos professores Sônia Maria de Stefano Piedade, Silvio Sandoval Zocchi, César Gonçalves de Lima e Roseli Aparecida Leandro, pelos ensinamentos e amizade.

Ao departamento de Ciências exatas da ESALQ-USP, por proporcionar formação de qualidade, com excelente estrutura física e corpo docentes.

Ao Prof^o Geraldo Veríssimo de Souza Barbosa, por ter me guiado nos primeiros passos dessa longa caminhada, pelo exemplo de pessoa e profissionalismo, por todo carinho, incentivo e conhecimento compartilhado, com certeza, o seu apoio e orientação foram fundamentais para escolha da minha formação profissional.

Ao Programa de Melhoramento Genético da Cana-de-açúcar da Universidade Federal de São Carlos - PMGCA-UFSCar pela bolsa concedida, sem a qual seria impossível a minha permanência em Piracicaba.

Aos Professores Hermann Hoffmann, Marcos Sanches, Antônio Bassinelo, que no momento mais difícil me deram apoio, amizade e carinho. Serei eternamente grata por ter sido acolhida com tanto amor, confiança e respeito.

A prof^a Monalisa Carneiro, que além das “caronas” dadas no meu deslocamento para Araras, mostrou-se amiga, companheira e carismática, com atitudes e profissionalismo.

A toda a equipe que forma o PMGCA-UFSCar que de uma forma especial tornam o ambiente de trabalho, acolhedor, amigável e prazeroso.

A todos os estagiários do PMGCA-UFSCar, pela amizade, dedicação e espírito de equipe.

A todos os companheiros do curso de estatística, que apesar da pouca convivência tenho uma admiração especial.

Um agradecimento especial eu reservo, a Erick Espinoza Nunez, o seu amor, compreensão, carinho e força, sempre me fizeram reagir e lembrar que vale apenas continuar na luta mesmo que os obstáculos pareçam invencíveis.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
1 INTRODUÇÃO	11
2 OBJETIVO	13
3 REVISÃO DE BIBLIOGRÁFICA	14
3.1 A Cana-de-açúcar	14
3.1.1 Importância para o Brasil	14
3.1.2 Melhoramento da Cana-de-açúcar no Brasil	15
3.2 Medidas Repetidas no Tempo	15
3.3 Análise de Perfis	21
3.3.1 Análise Multivariada de Perfis	22
3.3.2 Análise Univariada de Perfis	30
3.3.3 Teste de Hipótese sobre os parâmetros do modelo	31
3.3.4 Condição de Esfericidade da Matriz de Covariância	32
3.3.5 Teste de Mauchly	33
3.3.6 Correções do Número de Graus de Liberdade	34
3.3.7 Modelos Mistos	37
3.4 Estimação em Modelos Mistos	39
3.4.1 Estimação de β e predição de $\mathbf{u}$	39
3.4.2 Estimando os parâmetros de covariância em G e R	41
3.5 Estruturas da Matriz de Covariâncias	42
3.6 Seleção do Modelo e Teste Estatístico	45
4 MATERIAL E MÉTODOS	48
5 RESULTADOS E DISCUSSÃO	50
5.1 Análise Univariada	50
5.2 Análise Univariada e Correção dos Números de Graus de Liberdade	51
5.3 Análise Multivariada	53

5.4 Análise Via Modelo Misto	55
5.5 Análise Univariada	57
5.6 Análise Univariada e Correção dos Números de Graus de Liberdade	58
5.7 Análise Multivariada	59
5.8 Análise Via Modelo Misto	61
6 CONCLUSÕES	64
REFERÊNCIAS	65
ANEXOS	69

RESUMO

Análise de dados longitudinais em experimentos com cana-de-açúcar

Nesse trabalho foi abordada a situação em que observações de produtividade da cana-de-açúcar (TCH) foram tomadas na mesma unidade experimental em diferentes condições de avaliações (anos). Foram avaliados os perfis médios de resposta de 48 genótipos de cana-de-açúcar em dois experimentos: Experimento 1 e Experimento 2, durante três e cinco anos respectivamente, ambos com o delineamento de blocos ao acaso. Esse tipo de planejamento produz uma forma de relação entre as observações tomadas na mesma unidade experimental, portanto requer outras suposições, além das usuais, para que análise seja correta e os testes produzam resultados válidos. Para que as inferências sobre as médias de produtividade sejam válidas e seguras é necessário que o modelo da matriz de covariância dos dados seja apropriado. Diante disso, foram avaliadas três alternativas de análise para dados longitudinais (medidas repetidas no tempo), sendo utilizados portanto, o modelo univariado, conforme o planejamento do tipo “split-plot on time”, que impõe forte restrição quanto a matriz de variâncias-covariâncias; o modelo multivariado, que utiliza uma matriz de variâncias-covariâncias não-estruturada e o modelo misto, que possibilita a seleção de uma matriz que melhor representa os dados. Contudo, verificou-se que não houve diferença entre os resultados dos testes para as diferentes metodologias. Porém, é interessante a continuidade do estudo em relação ao modelo misto, pois devido a sua flexibilidade e precisão é possível obter estimativas mais seguras dos componentes de variância e prever os valores genotípicos, que por fim poderá proporcionar a previsão de produção de uma futura colheita para um determinado genótipo.

Palavras-chaves: Cana-de-açúcar; Medidas repetidas no tempo; Modelo univariado; Modelo multivariado; Modelo misto

ABSTRACT

Analysis of longitudinal data in experiments with sugar of cane

This work has been dealt with situation in which observations of productivity of sugar of cane (TCH) were taken in the same unit experimental in different condition of assessments (years). The response profiles average of 48 genotypes of sugar of cane were evaluated in two experiments: Experiment 1 and Experiment 2, for three and five years respectively, both with the randomized complete block design. This type of planning produces a form of relationship between the observations made in the same unit experimental therefore requires other assumptions, in addition to the usual, so that analysis is correct and the test results valid. To that inferences on the means of productivity are valid and safe it is necessary that the model of covariance matrix of the data is appropriate. Therefore, were evaluated three alternatives for analysis of longitudinal data (repeated measures over time), the univariate model as the planning of the split-plot on time which imposes strong restrictions on variances - covariances matrix, the multivariate model, which uses a non-structured variances - covariances matrix and mixed model, which they are enable the selection of a matrix that best represents the data. However, it was found that there was no difference between the results of tests for the different methodologies. But it is interesting the continuity of the study in relation to mixed model, because due to its flexibility and accuracy will be possible to obtain more reliable estimates of the variance components and predict the genotypic values, which ultimately could provide a prediction of production of a future harvest for a given genotype.

Keywords: Cane of sugar; Repeated measures over time; Univariate model; Multivariate model; Mixed model

LISTA DE FIGURAS

Figura 1 - Perfis médios de TCH dos 28 genótipos durante três anos	56
Figura 2 - Perfis médios de TCH dos 20 genótipos durante cinco anos	63

LISTA DE TABELAS

Tabela 1 - Critérios e Estatísticas Utilizadas para Testar a Hipótese Linear Geral . . .	29
Tabela 2 - Análise de Variância, Esperanças dos Quadrados Médios e Teste F .	31
Tabela 3 - Estrutura da Matriz de Covariâncias, Características . . .	44
Tabela 4 - Resultado da Análise Univariada Usando o proc GLM . .	51
Tabela 5 - Teste de Esfericidade	51
Tabela 6 - Análise Univariada e Correção dos Números de Graus de Liberdade para os Efeitos Intra-indivíduo	52
Tabela 7 - Resultado da Análise Multivariada Usando o proc GLM	54
Tabela 8 - Estrutura da Matriz de Covariâncias, Parâmetros Estimados e Critérios de In- formação de Akaike (AIC) e de Schwarz (BIC)	55
Tabela 9 - Teste para os Efeitos Fixos do Modelo com a Estrutura da Matriz de Co- variâncias do Tipo Simétrica Composta Heterogênea	56
Tabela 10 - Resultado da Análise Univariada Usando o proc GLM . .	57
Tabela 11 - Teste de Esfericidade	58
Tabela 12 - Análise Univariada e Correção dos Números de Graus de Liberdade para os Efeitos Intra-indivíduo	58
Tabela 13 - Resultado da Análise Multivariada Usando o proc GLM	61
Tabela 14 - Estrutura da Matriz de Covariâncias, Parâmetros Estimados e Critérios de In- formação de Akaike (AIC) e de Schwarz (BIC)	62
Tabela 15 - Teste para os Efeitos Fixos do Modelo com a Estrutura da Matriz de Co- variâncias do Tipo Auto-Regressiva Heterogênea	62
Tabela 16 - Produtividade de Cana-de-açúcar (TCH) dos Genótipos com três Cortes . .	70
Tabela 17 - Produtividade de Cana-de-açúcar (TCH) dos Genótipos com cinco Cortes .	71

1 INTRODUÇÃO

Muitos são os campos de estudo onde se efetuam várias observações sobre a mesma unidade experimental ao longo do tempo. Este tipo de planejamento é freqüentemente utilizado em diversas áreas de pesquisa, tais como: silvicultura, agricultura, principalmente com culturas perenes e semiperenes, medicina, psicologia e outras. Planejamento dessa natureza é bastante comum num programa de melhoramento genético de cana-de-açúcar, sendo a fase experimental, uma das etapas de grande importância para o pesquisador. Nessa etapa, são implantados vários experimentos objetivando avaliar e selecionar os melhores genótipos.

A classificação desses genótipos é feita através de um histórico de 11 a 13 anos do perfil agroindustrial dos clones-variedades. Essas informações estão contidas num banco de dados, e foram obtidas por meio de observações realizadas ao longo do tempo. Nessa fase, a ênfase maior é dada à produtividade agrícola que é considerada uma das principais características de interesse da indústria sucroalcooleira.

Durante a obtenção dos dados de campo, medidas dos caracteres de produção são tomadas ao longo do tempo nas mesmas unidades experimentais, ou seja, observam-se as produtividades dos genótipos (clones e variedades) em diferentes cortes, por exemplo, corte1 (cana-planta), corte2 (soca) e corte3 (ressoca), em anos sucessivos. Nesse contexto, o objetivo principal do pesquisador é conhecer o perfil produtivo dos genótipos ao longo do tempo.

Nesse tipo de ensaio, as unidades experimentais são sujeitas às medidas repetidas não aleatorizadas no tempo. Dessa forma, a análise de variância usual pode não ser válida, uma vez que na falta de aleatorização, os erros correspondentes às respectivas unidades experimentais podem ter uma matriz de variâncias-covariâncias que não é igual àquela exigida para que a análise usual de um delineamento seja válida, isto é, variâncias homogêneas.

Neste trabalho será abordada a situação em que observações de caracteres agroindustriais da cana-de-açúcar serão tomados na mesma unidade experimental, durante três e cinco anos sucessivos. Esse tipo de experimento produz uma forma de relação entre as observações obtidas nas diferentes ocasiões (tempos), portanto requer outras suposições, além das usuais, para que a análise seja correta e os testes produzam resultados válidos.

A análise de dados dessa natureza pode ser realizada através de um modelo univariado, de acordo com o planejamento do tipo “*split-plot on time*”, que impõe forte restrição

quanto à matriz de variâncias-covariâncias, como por meio de um modelo multivariado, que utiliza uma matriz de variâncias-covariâncias sem restrições, chamada não-estruturada. Uma outra forma de avaliar é através de modelos mistos lineares ou não-lineares, que possibilitam o uso de diferentes tipos de estrutura para as matrizes de variâncias-covariâncias.

Neste contexto, será utilizado o tempo como a variável representativa das ocasiões de observação, dando uma abordagem teórica das principais características, vantagens, desvantagens e restrições de cada uma das formas de análise de medidas repetidas, ressaltando as alternativas para análise das variáveis de produção da cana-de-açúcar, nos planejamentos em que as medidas são tomadas ao longo do tempo em um programa de melhoramento genético.

2 OBJETIVO

Os objetivos deste trabalho são:

- Comparar técnicas utilizadas para a análise de dados de medidas repetidas, utilizando o proc GLM e proc MIXED nos modelos univariado, multivariado e misto;
- Destacar novas alternativas para análise de dados longitudinais num programa de melhoramento genético de cana-de-açúcar;
- Propor uma técnica para análise experimentos com cana-de-açúcar, que seja mais eficiente e que melhore a precisão das estimativas de contrastes associados às diferenças entre os valores médios da resposta de diferentes cortes.

3 REVISÃO DE BIBLIOGRÁFICA

3.1 A Cana-de-açúcar

As primeiras notícias sobre cana-de-açúcar encontram-se nas escrituras mitológicas dos hindus, onde se declara que ela foi criada por Viswamitra, para o paraíso de Raja Ikkhakhu, e, quando esse paraíso foi destruído pelos demônios, foi permitido o uso da cana aos homens mortais da terra (MITRA, 1954) ¹. As Sagradas Escrituras também fazem referências sobre a cana, em Isaías, 43:24, e Jeremias, 6:20 (BÍBLIA SAGRADA, 1989).

Até o século 18 ela foi considerada como remédio e artigo de luxo. Sua crescente valorização como adoçante deve-se ao costume antigo de se adoçar chá, café e chocolate com mel. As suas qualidades como adoçante são relatadas mesmo antes de Cristo, quando o homem já a conhecia como “algo que produzia mel sem auxílio de abelhas”. Nessa época, foi constatado também que uma “bebida embriagante” podia ser fabricada a partir do seu caldo. Dessas observações, originou-se a crescente aceitação pela sociedade da época.

Atualmente, com a crise do petróleo e a emergência de novas fontes de energia, a cana-de-açúcar ganha destaque, dada as possibilidades reais do etanol no mercado de energias renováveis. Essa cultura tem produzido energia alternativa, substituindo, por vezes, o petróleo e é, sem dúvida alguma, a principal fonte capaz de tomar seu lugar.

3.1.1 Importância para o Brasil

A cana-de-açúcar é cultivada no Brasil desde a época do descobrimento, trazida por Martin Afonso de Souza em 1532. Os primeiros híbridos foram trazidos da Índia e de Java, na década de 20, seguidos por outros na década seguinte, e que substituíram com vantagem a Caiana e outras variedades antigas.

Atualmente, o Brasil é o maior produtor mundial de cana-de-açúcar e seus derivados, apresentando uma estimativa de 547,2 milhões de toneladas de cana-de-açúcar. A área plantada com a cultura alcançará 6,923 milhões de hectares, resultando em uma produtividade nacional de cerca de 79,034 toneladas por hectare (COMPANHIA NACIONAL

¹MITRA, S.M.N. Early history of sugarcane and sugar. **Journal of the Agricultural Society of Trinidad and Tobago**. Port of Spain, v. 54,n. 1, p. 39-41, 1954.

DE ABASTECIMENTO - CONAB, 2007).

3.1.2 Melhoramento da Cana-de-açúcar no Brasil

Nos programas de melhoramento de cana-de-açúcar, grande número de genótipos são avaliados todos os anos, gerando dados de produtividade com os quais procura-se indicar novas variedades para cada região produtora.

O sucesso dessa atividade econômica está atrelado principalmente aos trabalhos de pesquisa desenvolvidos nesses programas. A introdução da variedade melhorada é um dos fatores que mais contribui para essa liderança nacional, proporcionando o ótimo desempenho do setor canavieiro no país, através de cultivares com maior potencial produtivo (BARBOSA, 2005).

3.2 Medidas Repetidas no Tempo

O termo medidas repetidas é usado para designar medidas feitas no mesmo indivíduo ou na mesma unidade experimental em mais de uma ocasião (DIGGLE, 1988; CROWDER; HAND, 1990).

O estudo básico de medidas repetidas consiste de delineamentos experimentais completamente aleatorizados em que os tratamentos são alocados às unidades experimentais e os dados são coletados mais de uma vez em cada unidade experimental. Dessa forma, têm-se pelo menos dois fatores a serem estudados (tratamento e tempo), em que tratamento é o fator entre indivíduos e tempo, o fator intra-indivíduos (COSTA, 2003; ROSÁRIO et al., 2005).

Várias são as alternativas encontradas na literatura a respeito de planejamentos com medidas repetidas, sendo os mais comuns: “*split-plot*”, “*cross-over*” e planejamentos longitudinais. Os planejamentos do tipo “*split-plot*” (parcelas subdivididas) são comuns nas ciências agrárias, em que um único nível de um fator é alocado a uma parcela relativamente grande de terra e todos os níveis de um segundo fator, as subparcelas são aplicados. Nos planejamentos do tipo “*cross-over*”, as unidades experimentais recebem seqüências de tratamentos.

São freqüentes as situações em que o pesquisador toma medidas repetidas em diferentes tempos sobre uma unidade experimental, com o objetivo de verificar o comporta-

mento ao longo do tempo, caracterizando as medidas repetidas no tempo ou planejamentos longitudinais. Estes envolvem a observação de uma ou mais variáveis respostas em uma mesma unidade experimental em diversas condições de avaliação (tempo, distâncias de uma origem), são medidas correlacionadas e com variâncias não-homogêneas nos diversos tempos, em função do modo sistemático como as medidas são tomadas.

De acordo com Gill (1986); Lima (1996); Rosário (2003), entre as razões para a realização de experimentos dessa natureza, destacam-se a suspeita de que os efeitos dos tratamentos em uma seqüência de tempo se alteram, incorporando informações sobre a variação individual na análise, e o fato de que proporcionam condições adequadas para o controle de fatores acessórios que possam influenciar a resposta e melhoram, em geral, a precisão das estimativas de contrastes associados às diferenças entre as médias dos valores da resposta em diferentes ocasiões.

Rosário et al. (2005), em estudo de desempenho zootécnico de frangos de corte, avaliaram as variáveis zootécnicas (peso médio, ganho de peso, consumo de alimento e conversão alimentar) mensuradas semanalmente, inserindo o conceito de medidas repetidas. Os autores destacaram que o emprego de metodologias estatísticas apropriadas e eficientes à análise de dados dessa natureza aumenta a precisão na avaliação dos materiais genéticos de frangos de corte durante os testes de desempenho. Este fato é de extrema importância em um programa de melhoramento genético, pois possibilita a seleção dos materiais genéticos superiores ao longo dos ciclos de seleção.

Costa (2003) relata que estudos longitudinais constituem um caso especial dos estudos de medidas repetidas, que abrange os experimentos conhecidos como parcela subdividida e “*cross-over*”, assim como outros tipos de pesquisa em que cada unidade amostral é observada sob diferentes tratamentos e têm como objetivos:

- i) o estudo do comportamento da variável resposta ao longo do tempo e
- ii) a verificação da existência de dependência da variável resposta em relação às covariáveis.

Segundo o autor a característica distinta de estudos longitudinais é a dimensão ordenada com que os dados são coletados e o fato de que as observações repetidas para um indivíduo tendem a ser correlacionadas. Tal correlação pode ser modelada através de uma

estrutura de covariâncias dos dados observados sendo que, para outros tipos de dados, é usual assumir que os erros sejam independentes. Modelar uma estrutura de covariância apropriada é essencial para que as inferências sobre as médias sejam válidas.

Neste tipo de análise, a variável resposta pode ser contínua ou discreta, avaliada nas diferentes unidades experimentais (indivíduo, vasos, canteiros), que podem ser agrupadas segundo tratamentos ou fatores. Para cada unidade experimental obtêm-se diversas unidades de observação, que em conjunto, definem um perfil individual de respostas. Para cada tratamento (ou grupo) está associado um perfil médio de respostas que deve evidenciar o efeito do tratamento e o seu comportamento ao longo do tempo.

Os dados longitudinais são denominados regulares se o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do estudo. Além disso, se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais, teremos uma estrutura de dados balanceada. A ausência de observações perdidas caracteriza uma estrutura de dados completa.

De acordo com Xavier e Dias (2001) em um estudo de medidas repetidas, a estrutura de parcelas subdivididas, é caracterizada quando se aplicam às parcelas os níveis do fator A e nos quais se tomam medidas repetidas, em ocasiões sucessivas, sob a mesma parcela, admitindo-se que essas medidas, tomadas em ocasiões distintas, têm variâncias homogêneas e são igualmente correlacionadas.

Os autores ainda destacam que em estudos de medidas repetidas no tempo, em um delineamento no esquema de parcelas subdivididas, por exemplo, os níveis desse tempo não podem ser aleatorizados para seus intervalos. Nesse caso, a validade da análise de variância usual é duvidosa, pois na falta de aleatorização os erros correspondentes às respectivas unidades experimentais podem ter uma matriz de covariâncias que não é igual àquela exigida para que a análise usual de um delineamento seja válida, isto é, variâncias homogêneas.

Conforme Littell et al. (1998); Xavier (2000); Costa (2003); Rosário (2003), há semelhanças entre os experimentos com medidas repetidas e aqueles em parcelas subdivididas, em que os fatores tratamento e tempo correspondem, respectivamente, à parcela e subparcela no experimento em parcelas subdivididas. A diferença entre eles é que, nos experimentos em

parcelas subdivididas, os níveis da subparcela são aleatoriamente atribuídos às unidades de subparcela dentro das unidades de parcelas. Conseqüentemente, as respostas de diferentes subparcelas, na mesma parcela, são igualmente correlacionadas umas com as outras. Já em experimentos com medidas repetidas, as respostas de tempos mais próximos são, em geral, mais fortemente correlacionadas do que as de tempos mais distantes.

Rezende et al. (1999); Rosário et al. (2005) acrescentam que os experimentos com medidas repetidas, quando analisados sob o enfoque de parcela subdividida, violam duas pressuposições básicas requeridas pela análise de variância: a falta de casualização entre os tratamentos e as épocas de avaliação (tempo) e a dependência de erros, devido as medidas serem mensuradas nas mesmas unidades experimentais ao longo do tempo, ocasionando a correlação entre os dados.

Gill (1986); Riboldi et al. (1996) alertam que uma conseqüência imediata de se ignorar a correlação entre os dados mensurados em tempos adjacentes é que a significância aparente da diferença entre as médias dos tratamentos é grosseiramente exagerada e a sensibilidade dos testes para interação é seriamente reduzida. Quando a correlação de erros é ignorada, as inferências podem ou não ser distorcidas, conforme o grau de homogeneidade das variâncias e covariâncias dos dados nas diferentes épocas.

Fernandez (1991) também salienta o problema de que quando o experimento é sistematicamente arranjado, sem aleatorização, a análise de um experimento de medidas repetidas com delineamento de parcelas subdivididas pode inflacionar a probabilidade de falsamente rejeitar a hipótese nula (erro tipo I).

Huynh e Feldt (1970) mostraram que, em um delineamento de parcelas subdivididas com medidas repetidas no tempo, o teste F com relação à parcela tem distribuição F exata, mas com relação a subparcela, só terá distribuição F exata se a matriz de covariâncias satisfizer certa pressuposição, além das citadas anteriormente. Maiores detalhes sobre essa pressuposição serão apresentadas no decorrer do texto.

Outros autores como, Milliken e Johnson (1992), advertem que, tais pressuposições nem sempre são apropriadas para um delineamento em parcelas subdivididas com medidas repetidas no tempo, sendo, porém, uma análise correta quando realizada sob suposições mais gerais. Essas suposições mais gerais requerem certa forma para a matriz de

covariâncias dos erros denotada por Σ .

Xavier (2000) também destaca que uma condição suficiente para que o teste F da análise de variância usual, em nível de subparcela, para o fator tempos e interação tempos $\times$ tratamentos, seja válido, é que a matriz de covariâncias tenha uma forma chamada de simétrica composta, que ocorre quando a matriz de covariâncias Σ puder ser expressa, por exemplo, como:

$$\Sigma = \begin{bmatrix} (\sigma_\gamma^2 + \sigma_\varepsilon^2) & \sigma_\gamma^2 & \sigma_\gamma^2 & \sigma_\gamma^2 \\ \sigma_\gamma^2 & (\sigma_\gamma^2 + \sigma_\varepsilon^2) & \sigma_\gamma^2 & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\gamma^2 & (\sigma_\gamma^2 + \sigma_\varepsilon^2) & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \sigma_\gamma^2 & (\sigma_\gamma^2 + \sigma_\varepsilon^2) \end{bmatrix}$$

em que σ_ε^2 é a variância da subparcela (intra-indivíduos), σ_γ^2 é a variância da parcela (entre indivíduos).

A autora comenta que a condição de simetria composta implica que a variável aleatória seja igualmente correlacionada e tenha variâncias iguais, considerando as diferentes ocasiões. Nesse caso, diz-se que a matriz de covariâncias é do tipo uniforme, ou segue o padrão de uniformidade, ou tem a forma de simetria composta (a variância das respostas em qualquer um dos tempos é igual a $\sigma_\gamma^2 + \sigma_\varepsilon^2$ e a covariância entre dois tempos quaisquer é igual a σ_γ^2).

A análise feita através desse modelo corresponde à análise de um experimento em parcelas subdivididas ou “*split-plot*” em que as causas de variação entre indivíduos (tratamento) são agrupadas separadamente daquelas que fazem parte da variação intra-indivíduos (tempo e interação tratamento $\times$ tempo).

Apesar das facilidades de obtenção e de interpretação dos resultados dos testes das hipóteses, a aplicação dessa abordagem não é recomendada para a análise de dados longitudinais, pois considerando o modo sistemático como são feitas as observações nas mesmas unidades experimentais, não se espera que a matriz Σ seja do tipo uniforme.

Huynh e Feldt (1970); Rouanet e Lépine (1970) mostraram que uma condição suficiente e necessária para que as estatísticas dos testes de hipótese envolvendo comparações intra-indivíduos tenham distribuição F exata é que a matriz de covariâncias Σ satisfaça a

condição de esfericidade ou circularidade. Os autores descreveram uma condição mais geral da forma de Σ , a qual foi denominada de HUYNH - FELDT (H-F) e especifica que os elementos da matriz de covariâncias Σ sejam expressos, para um $\lambda > 0$, como:

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

em que λ é a diferença entre a média das variâncias e a média das covariâncias.

Xavier e Dias (2001) também salientam que a condição H-F é necessária e suficiente para que o teste F da análise de variância usual, no esquema de delineamento de parcelas subdivididas para análise de dados com medidas repetidas no tempo, seja válido. Essa condição é equivalente a especificar que variâncias da diferença entre pares de erros sejam todas iguais, e se as variâncias são todas iguais então a condição é equivalente a de simetria composta. Isso pode ser verificado calculando-se todas as variâncias das diferenças dos possíveis pares de erros e mostrando que

$$\sigma_{Y_j - Y_{j'}}^2 = \sigma_{Y_j}^2 - \sigma_{j'}^2 - 2\sigma_{jj'}^2 = 2\lambda,$$

é constante para todo j e j' ($j \neq j'$), isto é, desde que $\sigma_{Y_j - Y_{j'}}^2$ seja igual a uma constante para todo j e j' ($j \neq j'$). A matriz Σ é dita do tipo H-F, e se as variâncias são todas iguais então a condição é equivalente à de simétrica composta.

As matrizes de covariâncias Σ , na forma da simetria composta e erros independentes, são casos especiais da condição de H-F.

A maioria dos autores questiona que um problema com relação à validade dos testes surge quando se têm estruturas da matriz de covariâncias diferentes das estruturas de simetria composta, erros independentes e da condição de H-F, levando a testes F não exatos.

Meredith e Stehman (1991) verificaram que a violação da condição de H-F leva a testes muito liberais para os fatores da subparcela (para tempos e para a interação tratamentos $\times$ tempos).

Malheiros (2004), num trabalho sobre a precisão de teste F em experimentos com medidas repetidas no tempo, destaca que alguns autores, como Box (1954 I, II), Geisser e

Greenhouse (1958)² e Huynh e Feldt (1976) sugerem algumas correções nos graus de liberdade para o fator tempos e para a interação entre tratamentos e tempos, possibilitando que o teste possa ser usado, de forma aproximada, mesmo que a condição de esfericidade da matriz Σ não seja satisfeita. Nesse caso, o problema que surge é que pouco se sabe sobre a precisão dos testes F quando essas correções são realizadas, e se a precisão depende da estrutura da matriz Σ e dos dados serem balanceados ou não. Littell et al. (1998) sugerem o uso da correção de Greenhouse e Geisser.

Conhecendo a problemática, Mauchly (1940)³ propôs um teste chamado teste de esfericidade. Através dele é possível verificar se a matriz de covariâncias atende a condição de H-F. Esse teste consiste em verificar se uma população multivariada apresenta variâncias iguais e correlações nulas. Esse teste está implementado, por exemplo, no proc GLM do SAS e pode ser realizado utilizando-se o comando “repeated” com a opção “printe”.

Xavier e Dias (2001) também salientam que, caso a condição de H-F para a matriz Σ de covariâncias não seja satisfeita, uma alternativa seria a análise multivariada, também conhecida como análise de perfis, que adota uma hipótese mais geral sobre a estrutura da matriz de covariâncias. Outra possibilidade seria utilizar análise univariada no esquema de parcelas subdivididas para análise de dados de medidas repetidas, realizando o ajuste do número de graus de liberdade do teste F para o fator da subparcela.

A seleção do melhor modelo de estrutura da matriz pode, então, ser feita utilizando o Critério de Informação de Akaike (AIC), Critério de Informação de Schwarz (BIC) ou através de um teste de razão de máxima verossimilhança.

3.3 Análise de Perfis

Segundo Aubin (1984), a Análise de Perfis é uma técnica clássica de análise de dados longitudinais que pode ser realizada utilizando-se técnicas multivariadas ou univariadas 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

²GEISSER, J.; GREENHOUSE, S.W. An extension of Box’s results on the use of the F distribution in multivariate analysis. **Annals of the Mathematical Statistics**, v.29, p.855-891, 1958.

³MAUCHLY, J.W. Significance test for sphericity of a normal n-variate distribution. **Annals of Mathematical Statistics**, v. 11, p.204-209, 1940.

experimentais, ou tratamentos.

Lima (1996) também afirma que a Análise de Perfis é realizada com o objetivo de estudar o efeito dos tratamentos nas diversas ocasiões de avaliação, em outras palavras, consiste em testar hipótese sobre os perfis médios de respostas dos tratamentos, isto é, sobre os valores médios da variável resposta nas diferentes condições de observação (tempo). Pode ser realizada utilizando-se técnicas de análise multivariadas ou univariadas e a escolha por uma dessas técnicas depende das suposições que podemos admitir como verdadeiras para o conjunto de dados em estudo. Uma ferramenta muito útil nesses estudos é o proc GLM do SAS.

3.3.1 Análise Multivariada de Perfis

A análise de perfis por meio de técnicas multivariadas consiste numa alternativa para análise de medidas repetidas no tempo. Essa técnica é bastante discutida na literatura e dentre os autores podem-se citar, Morrison (1990), Andrade e Singer (1986), Lima (1996) e outros.

Segundo Singer e Andrade (1986) a análise multivariada de perfis também chamada de análise de variância multivariada, é uma solução natural para dados de medidas repetidas, pois pode apresentar menor poder em seus testes e às vezes indicar diferenças significativas que realmente, não existem. Porém, para minimizar esses riscos deve-se ter a garantia de que os erros tenham distribuição normal multivariada.

A análise multivariada não impõe nenhuma restrição ou condição para a matriz de covariância dos efeitos. Meredith e Stehrnan (1991) também ressaltam que não há suposição sobre a estrutura da matriz de covariâncias, sendo, por isso, uma solução natural para dados de medidas repetidas.

Lima (1996) salienta que este enfoque multivariado permite a adoção de um modelo bastante geral para representar a estrutura de covariâncias, admitindo que as variâncias das respostas em cada tempo e as covariâncias das respostas entre tempos distintos sejam diferentes. O autor destaca como principais desvantagens a necessidade de perfis individuais de resposta completos e o baixo poder dos testes de hipóteses, devido à excessiva parametrização.

Vonesh e Chinchilli (1997) acrescentam que, geralmente, as técnicas usuais

impõem suposições de que todas observações sejam independentes, mas essa suposição não é adequada para dados de medidas repetidas onde as observações feitas no mesmo indivíduo usualmente são correlacionadas. Os autores sugerem, para um experimento com delineamento de parcelas subdivididas, com medidas repetidas no tempo, o seguinte modelo:

$$y_{ijk} = \mu + \beta_i + \tau_j + \gamma_k + (\tau\gamma)_{jk} + e_{ijk}$$

em que y_{ijk} é o valor observado para a variável resposta no k -ésimo tempo para o j -ésimo tratamento no i -ésimo bloco, μ é uma constante comum a todas as observações, β_i é o efeito do i -ésimo bloco, τ_j é o efeito do j -ésimo tratamento, γ_k é o efeito do k -ésimo tempo observado, $(\tau\gamma)_{jk}$ é o efeito da interação entre o j -ésimo tratamento com o k -ésimo tempo e e_{ijk} é o erro aleatório associado às observações do k -ésimo tempo para j -ésimo tratamento no i -ésimo bloco (variação do acaso sobre as observações), supostos homocedásticos, independentes e normalmente distribuídos, para $i = 1, \dots, b$, $j = 1, \dots, g$, $k = 1, \dots, t$.

Nesse modelo, observa-se que o erro da parcela não está incluído. Para que o modelo tenha posto completo, é necessário impor algumas restrições, conforme segue:

$$\sum_{i=1}^b \beta_i = 0; \quad \sum_{j=1}^g \tau_j = 0; \quad \sum_{k=1}^t \gamma_k = 0; \quad \sum_{j=1}^g (\tau\gamma)_{jk} = 0, \text{ para } k = 1, \dots, t \text{ e } \sum_{k=1}^t (\tau\gamma)_{jk} = 0,$$

para $j = 1, \dots, g$ e com o vetor aleatório

$$e_{ij} = [e_{ij1} \cdots e_{ijt}]' \sim i.i.d.N_t(0, \Sigma)$$

em que Σ é uma matriz $t \times t$, positiva definida com uma estrutura geral.

O modelo usado para análise multivariada de perfis pode ser representado, matricialmente, na forma usual da Análise Multivariada de Variância (MANOVA), isto é:

$$\mathbf{Y} = \mathbf{X}\beta + \Psi$$

em que $\mathbf{Y}_{(n \times t)} = [\mathbf{y}_{11}, \dots, \mathbf{y}_{gn_g}]'$ é a matriz dos dados observados de t respostas para os n indivíduos, $\mathbf{y}_{ij} = [y_{ij1}, y_{ij2}, \dots, y_{ijt}]'$ é o perfil de resposta da unidade experimental (i, j) , $\mathbf{X}_{(gb \times (g+b+1))}$ é a matriz de delineamento conhecida. Essa matriz corresponde aos valores da variável explanatória e das variáveis “dummy” associadas com a classificação das variáveis, $\beta_{((g+b+1) \times t)}$ é a matriz de parâmetros dos efeitos fixos desconhecidos e $\Psi_{(gb \times t)}$ é a matriz do erro experimental.

Sendo:

$$\mathbf{X}_{(gb \times (g+b+1))} = [X_1 : X_2 : X_3], \text{ com } X_1 = [1_{gb}], X_2 = \begin{bmatrix} \mathbf{I}_b \\ \mathbf{I}_b \\ \vdots \\ \mathbf{I}_b \end{bmatrix} \text{ e } X_3 = \begin{bmatrix} 1_b & 0 & \dots & 0 \\ 0 & 1_b & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1_b \end{bmatrix}$$

com 1_{gb} um vetor com gb elementos iguais a um, $\mathbf{I}_{(b \times b)}$ é uma matriz identidade e $1_{(b \times 1)}$ é um vetor de uns e a matriz dos parâmetros:

$$\beta_{((g+b+1) \times t)} = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_t \\ \beta_{11} & \beta_{12} & \dots & \beta_{1t} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{b1} & \beta_{b2} & \dots & \beta_{bt} \\ \tau_{11} & \tau_{12} & \dots & \tau_{1t} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{g1} & \tau_{g2} & \dots & \tau_{gt} \end{bmatrix}$$

para

$$\mu = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_t \end{bmatrix}', \beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1t} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{b1} & \beta_{b2} & \dots & \beta_{bt} \end{bmatrix}, \tau = \begin{bmatrix} \tau_{11} & \tau_{12} & \dots & \tau_{1t} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{g1} & \tau_{g2} & \dots & \tau_{gt} \end{bmatrix}$$

Neste modelo, a matriz do erro experimental é

$$\Psi = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{gb} \end{bmatrix}', \text{ com } \varepsilon_{ij} = \begin{bmatrix} \varepsilon_{ij1} & \varepsilon_{ij2} & \dots & \varepsilon_{ijt} \end{bmatrix}'$$

que é o vetor de erro experimental associado ao i -ésimo indivíduo, sendo que as linhas ε_{ij} da matriz Ψ são supostas independentes e identicamente distribuídas com distribuição normal

t -variada com média 0 e matriz de covariâncias comum, Σ . Assim, enquanto as linhas de Ψ são independentes, os elementos das suas colunas podem ser correlacionados. Dessa forma, para o modelo usual linear multivariado, assume-se que:

$$\mathbf{E}[\varepsilon_{ij}] = 0 \text{ e } \mathbf{Var}[\varepsilon_{ij}] = \Sigma, \quad i = 1, 2, \dots, n.$$

Logo, na forma matricial tem-se que

$$\Psi_{(gb \times t)} = \begin{bmatrix} \varepsilon_{111} & \varepsilon_{112} & \dots & \varepsilon_{11t} \\ \varepsilon_{121} & \varepsilon_{122} & \dots & \varepsilon_{12t} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{1b1} & \varepsilon_{1b2} & \dots & \varepsilon_{1bt} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{g11} & \varepsilon_{g12} & \dots & \varepsilon_{g1t} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{gb1} & \varepsilon_{gb2} & \dots & \varepsilon_{gbt} \end{bmatrix}.$$

Por extensão, supõe-se que os perfis de respostas $\mathbf{y}_{ij}$ são independentes e apresentam distribuições Normais t -variadas e as matrizes de covariâncias correspondentes sejam todas iguais. Assim, a média e a variância de cada perfil podem ser representadas da seguinte forma

$$E[\mathbf{y}_{ij}] = \mathbf{X}\beta, \text{ com } \mathbf{Var}[\mathbf{y}_{ij}] = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1t} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1t} & \sigma_{2t} & \dots & \sigma_t^2 \end{bmatrix}.$$

Portanto, $\mathbf{y}_{ij} \sim i.i.d.N_t(\mathbf{X}\beta; \Sigma)$.

Nesse modelo, a matriz de variâncias e covariâncias é chamada não estruturada ou completamente parametrizada e tem $t(t+1)/2$ parâmetros, conforme mencionado anteriormente. Também é possível notar que é preciso admitir que todos os grupos (tratamentos) devem ter a mesma matriz de variâncias e covariâncias, Σ .

Nesse estudo, as hipóteses de interesse são:

i) $\mathbf{H}_{0I}$: Os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos, isto é, a interação entre os fatores tratamentos e tempos é nula. Na forma matricial, em termos dos parâmetros do modelo, a hipótese $\mathbf{H}_{0I}$, pode ser expressa na forma:

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

Essa hipótese é fundamental, pois sua validade pode determinar a forma das outras duas hipóteses de interesse. Quando o paralelismo dos perfis médios de resposta é aceitável, devem-se estudar as seguintes hipóteses:

ii) $\mathbf{H}_{0G}$: Os perfis médios de resposta correspondentes aos diversos tratamentos são coincidentes, isto é, não existe efeito desse(s) fator(es). Em termos dos parâmetros do modelo, a hipótese $\mathbf{H}_{0G}$, pode ser expressa como:

$$\mathbf{H}_{0G} = \sum_{k=1}^t \mu_{1k} = \sum_{k=1}^t \mu_{2k} = \dots = \sum_{k=1}^t \mu_{gk}$$

ou na forma matricial:

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

iii) $\mathbf{H}_{0T}$: Os perfis médios de resposta correspondentes aos diversos tratamentos são paralelos ao eixo das abscissas, isto é, não existe efeito do fator tempos. Em termos dos parâmetros do modelo, essa hipótese, pode ser expressa como:

$$\mathbf{H}_{0T} = \sum_{i=1}^g \mu_{i1} = \sum_{i=1}^g \mu_{i2} = \dots = \sum_{i=1}^g \mu_{it}$$

ou na forma matricial:

$$\mathbf{H}_{0T} = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{g2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{gt} \end{bmatrix}$$

Para unificar os resultados estatísticos, podem-se expressar todas as hipóteses apresentadas na forma da hipótese linear geral:

$$\mathbf{H} : \mathbf{G}\beta\mathbf{T} = \mathbf{0}$$

em que $\mathbf{G}_{((g-1) \times (g+b+1))}$ e $\mathbf{T}_{(t \times (t-1))}$ são matrizes de constantes conhecidas com posto g e t , respectivamente. Pode-se observar que a matriz $\mathbf{G}$ é responsável por comparações entre tratamentos (linhas da matriz β), ou seja, entre os genótipos, e a matriz $\mathbf{T}$, por comparações entre as ocasiões de observação (colunas da matriz β), nesse caso, o tempo. Através da hipótese linear geral, é possível obter as seguintes correspondências:

$$\mathbf{H}_{0I} : \mathbf{G}_{1(g-1) \times (g+b+1)} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 1 & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

e

$$\mathbf{T}_{1(t \times (t-1))} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{bmatrix}$$

$$\mathbf{H}_{0G} : \mathbf{G}_1 \text{ e } \mathbf{T}_2 = \mathbf{1}_t$$

$$\mathbf{H}_{0T} : \mathbf{G}_2 = \mathbf{1}'_g \text{ e } \mathbf{T}_1$$

em que $\mathbf{1}_t$ e $\mathbf{1}'_g$ são vetores de uns com dimensões t e g , respectivamente, e $\mathbf{G}_1$ e $\mathbf{T}_1$, já definidos anteriormente. Convém salientar que essas representações não são únicas.

Segundo Ferreira (1996), testes para hipótese linear geral são realizados considerando as magnitudes das somas de quadrados e produtos de tratamento e resíduo pela variância generalizada. Esses testes, de acordo com Singer e Andrade (1986) podem ser obtidos através de diversos critérios; em geral as estatísticas de teste correspondente são funções das raízes características da matriz HE^- , sendo que

$$H = T'((X'X)^-X'Y)G'[G(X'X)^-G']^-G((X'X)^-X'Y)T$$

é a matriz de somas de quadrados e produtos cruzados devido à hipótese nula, e

$$E = T'Y'[I - X(X'X)^-X']^-YT$$

é a matriz de somas de quadrados e produtos cruzados devido ao erro.

Dentre as várias estatísticas de teste disponíveis, tem-se: Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Maior Raiz Característica de Roy, sendo que as mais difundidas são obtidas através dos princípios da união-intersecção de Roy e da razão de verossimilhança de Wilks.

Muitos autores recomendam utilizar o critério de Wilks como referência, por se tratar de um teste baseado na razão de verossimilhança. Outros recomendam que a hipótese nula deva ser rejeitada se pelo menos três dos quatro critérios forem significativos a um nível nominal de significância previamente adotado.

Definindo $\theta_i = \lambda_i(1 + \lambda_i)^{-1}$, em que λ_i é a i -ésima raiz característica de HE^- , essas estatísticas são mostradas na Tabela 1.

Tabela 1 - Critérios e Estatísticas Utilizadas para Testar a Hipótese Linear Geral

Critério	Estatística
Lambda de Wilks	$\Lambda = \prod_{i=1}^S (1 - \theta_i)$
Raiz máxima de Roy	$\theta_S = \max(\theta_i)$
Traço de Hotelling-Lawley	$T = \sum_{i=1}^S \theta_i (1 - \theta_i)^{-1}$
Traço de Pillai	$P = \sum_{i=1}^S \theta_i$

Obs: $S = \min(t - 1, g - 1)$, onde t são os tempos e g os tratamentos

As distribuições exatas dessas estatísticas, sob a hipótese nula, dependem unicamente dos parâmetros $m_1 = (|(t - 1) - (g - 1)| - 1)/2$ e $m_2 = (n - g - (t - 1) - 1)/2$ e S , já definido. Detalhes sobre essas estatísticas podem ser encontradas em Timm (1975) ou Morrison (1990), e também tabelas com os valores críticos da estatística de Roy.

Xavier (2000) ainda destaca que, esses testes não requerem a condição de H-F, pois são baseados em matriz de covariância sem estrutura. As estimativas da matriz de covariância são obtidas pelo método dos momentos.

Porém, aplicação da técnica de análise multivariada de perfis, segundo Lima (1996), apresenta algumas restrições, sendo indicada para situações em que o número total de unidades experimentais é maior do que o número de ocasiões (tempo), $n > t$. Meredith e Stehman (1991) advertem que uma desvantagem dessa análise é a falta de poder para estimar os parâmetros da matriz de covariâncias, quando t é grande e n pequeno. Nos casos para os quais isso ocorre, Singer e Andrade (1986) afirmam que é necessário impor certas restrições na matriz de covariância Σ .

Outras desvantagens dessa técnica são:

i) Necessidade de perfis de dados completos (na perda de uma ou mais observações para um mesmo indivíduo; todo o perfil de respostas desse indivíduo é excluído da análise);

ii) As diferentes estatísticas de teste podem levar a conclusões diferentes. Nesse caso, uma sugestão é assumir como verdadeiro, o resultado mais comum entre as estatísticas.

Usualmente, sobre a condição H-F, as técnicas de análise univariadas de perfil são mais poderosas do que as técnicas multivariadas, pois a primeira apresenta uma maior probabilidade de detectar efeitos significativos quando esses realmente existem.

3.3.2 Análise Univariada de Perfis

O modelo para análise univariada de perfil, corresponde ao adotado na análise de experimentos em parcelas subdivididas “split-plot”, em que as causas de variação entre indivíduos são agrupadas separadamente daquelas de variação intra-indivíduos. Na análise de variância adotou-se o seguinte modelo proposto por Vonesh e Chinchilli (1997):

$$y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \gamma_k + (\tau\gamma)_{jk} + e_{ijk}$$

em que y_{ijk} é o valor observado para a variável resposta no k -ésimo tempo para o j -ésimo tratamento no i -ésimo bloco, μ é uma constante comum a todas as observações, β_i é o efeito do i -ésimo bloco, τ_j é o efeito do j -ésimo tratamento, γ_k é o efeito do k -ésimo tempo observado, $(\beta\tau)_{ij}$ é o efeito aleatório da “interação” do i -ésimo bloco com o j -ésimo tratamento (erro associado às parcelas), $(\tau\gamma)_{jk}$ é o efeito da interação entre o j -ésimo tratamento com o k -ésimo tempo e e_{ijk} é o erro aleatório associado às observações do k -ésimo tempo para j -ésimo tratamento no i -ésimo bloco (variação do acaso sobre as observações), supostos homocedásticos, independentes e normalmente distribuídos, para $i = 1, \dots, b$, $j = 1, \dots, g$ e $k = 1, \dots, t$.

No modelo univariado no esquema de parcelas subdivididas, são feitas pressuposições de que, tanto o erro ao nível de parcela, que engloba o fator de tratamentos $\times$ blocos, como o erro da subparcela, onde são alocados o fator tempo e interação tempos $\times$ tratamentos, tenham distribuição normal, sejam independentes e identicamente distribuídos, com variâncias constantes, cujas pressuposições são as mesmas feitas para uma análise usual. O erro da parcela também é conhecido como erro entre indivíduos, e o erro da subparcela como intra-indivíduos.

O quadro da análise da variância e as expressões para o teste F para os fatores de interesse do modelo univariado está representado na Tabela 2.

Tabela 2 - Análise de Variância, Esperanças dos Quadrados Médios e Teste F

Causas de Variação	G.L.	Q.M.	E(Q.M.)	F
<i>Bl.</i> (B)	$(b - 1)$	QM_1	$\sigma^2 + t\sigma_{bg}^2 + gt\sigma_b^2$	$F_1 = QM_1/QM_3$
<i>Trat.</i> (G)	$(g - 1)$	QM_2	$\sigma^2 + t\sigma_{bg}^2 + bt\phi_g$	$F_2 = QM_2/QM_3$
<i>Res.</i> (BG)	$(b - 1)(g - 1)$	QM_3	$\sigma^2 + t\sigma_{bg}^2$	
<i>Tempo</i> (T)	$(t - 1)$	QM_4	$\sigma^2 + gb\phi_t$	$F_3 = QM_4/QM_6$
<i>Trat. x Tempo</i> (GT)	$(g - 1)(t - 1)$	QM_5	$\sigma^2 + b\phi_{gt}$	$F_4 = QM_5/QM_6$
<i>Res.</i> (BGT)	$g(b - 1)(t - 1)$	QM_6	σ^2	
<i>Total Cor.</i>	$gbt - 1$			

Note que:

$$\phi_g = \frac{\sum_j \tau_j^2}{g - 1}, \quad \phi_t = \frac{\sum_k \gamma_k^2}{t - 1}, \quad \phi_{gt} = \frac{\sum_{jk} (\tau\gamma)_{jk}^2}{(g - 1)(t - 1)}.$$

3.3.3 Teste de Hipótese sobre os parâmetros do modelo

Através do método de análise univariada “split-plot”, será testado se existe diferença entre tratamentos, entre as condições de avaliação (tempos) e entre as interações (tratamento $\times$ tempo). Para isso, serão testadas as seguintes hipóteses de interesse:

1) Para efeito de tratamentos

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \dots = \tau_g = 0$$

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

2) Para efeito de tempo

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \dots = \gamma_t = 0$$

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

3) Para efeito da interação (tratamento $\times$ tempo)

$$H_0 : (\tau\gamma)_{11} = (\tau\gamma)_{12} = (\tau\gamma)_{13} = \dots = (\tau\gamma)_{gt} = 0$$

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

As hipóteses nulas, associadas aos efeitos principais de tratamento, às condições de avaliação (tempos) e as interações (tratamento $\times$ tempo), podem ser facilmente testadas através das estatísticas, F_2, F_3 e F_4 , apresentadas na Tabela 2, que, sob as respectivas hipóteses nulas, seguem distribuições exatas F centrais com $(g-1)$ e $(b-1)(g-1)$; $(t-1)$ e $g(b-1)(t-1)$; $(g-1)(t-1)$ e $g(b-1)(t-1)$ graus de liberdade, respectivamente.

É comum adotar o nível mínimo de significância de 0.05 para testar a hipótese de nulidade (H_0). Então, sempre que o valor da probabilidade do teste F for menor ou igual a 0.05, rejeita-se H_0 , pois há diferença significativa entre pelo menos dois níveis dos fatores. É interessante destacar que, para interpretar os resultados dos testes F, deve-se começar com as interações duplas, pois quando a interação não for significativa, o passo seguinte será o estudo dos efeitos principais do modelo.

Conforme comentado anteriormente, em estudo de planejamento longitudinal, não se espera que a matriz Σ , siga o padrão de uniformidade. E mesmo diante da facilidade de obtenção e interpretação das estatísticas F's, a aplicação direta dessa abordagem, não é recomendada para análise de dados longitudinais. Embora, fatores como a simplicidade na obtenção, interpretação e a ampla divulgação dos modelos univariados, contribuam substancialmente para o uso indiscriminado dessas estatísticas por muitos pesquisadores.

Dessa forma, vários autores sugerem a correção no número de graus de liberdade da estatística F, a fim de conseguir o padrão de uniformidade de Σ .

3.3.4 Condição de Esfericidade da Matriz de Covariância

O padrão de uniformidade para Σ é uma condição suficiente mas não necessária para que F_3 e F_4 tenham distribuição exata F central. Em trabalhos independentes, Rouanet e Lépine (1970) e Huynh e Feldt (1970), demonstraram que a condição necessária e suficiente para que as estatísticas F_3 e F_4 sejam válidas é que os elementos $\sigma_{kk'}$, k e $k' = 1, 2, \dots, t$ de Σ , satisfaçam:

$$\sigma_{kk'} = \begin{cases} a_k + a_{k'} + \lambda, & \text{se } k = k' \\ a_k + a_{k'}, & \text{se } k \neq k' \end{cases}$$

com $a_k, a_{k'}$ e $\lambda > 0$ constantes. Desta forma, a matriz de covariâncias Σ , satisfaz à condição de circularidade ou esfericidade. O que equivale a

$$\mathbf{C}_{(t-1) \times t} \Sigma_{(t \times t)} \mathbf{C}'_{t \times (t-1)} = \lambda \mathbf{I}_{(t-1)}, \quad (1)$$

em que $\mathbf{C}$ é a matriz dos coeficientes dos contrastes ortonormais que representa o total de hipótese nula, Σ é a matriz de covariâncias, λ é um escalar maior que zero e $\mathbf{I}$ é a matriz identidade $(t-1) \times (t-1)$.

Alguns autores argumentam que, apesar dessa estrutura da matriz Σ não ser tão restritiva quanto à estrutura uniforme, ela também não é adequada para representar a estrutura de covariância de planejamento longitudinal, pois espera-se que a correlação entre duas observações tomadas em tempos mais próximos seja maior do que quando tomadas em tempos mais distantes.

3.3.5 Teste de Mauchly

Para verificar a validade da condição de esfericidade da matriz de covariâncias, Mauchly (1940)³ citado pelos autores, Morrison (1990); Kuehl (1994) e Kirk (1995); Xavier (2000), desenvolveu um teste que verifica se uma população normal multivariada apresenta variâncias iguais e correlações nulas. Nesse caso, uma população será chamada de “esférica”, quando tiver essa simetria.

A estatística de teste formulada por Mauchly (1940)³ na verificação da condição de esfericidade 1 é

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

A descrição do teste de esfericidade apresentada por Kuehl (1994) e Kirk (1995), foi relatada da seguinte forma: admitiu que s_{ij} é o elemento da i -ésima linha e j -ésima coluna da matriz de covariância amostral $\mathbf{S}_{(t \times t)}$, para o erro intra-indivíduos, com ν graus de liberdade. Nas t medidas repetidas, foram escolhidos $(t-1)$ contrastes ortogonais normalizados e sabendo que as linhas da matriz $\mathbf{C}_{(t-1) \times t}$ são contrastes ortogonais normalizados nas t medidas repetidas, calcula-se então, a matriz $\mathbf{CSC}'_{(t-1) \times (t-1)}$.

Para melhorar a acurácia dessa aproximação pela distribuição Qui-quadrado,

foi definido o seguinte fator escalar

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

Então, a hipótese de que a matriz de covariâncias satisfaz à condição de esfericidade, ou seja, hipótese $\mathbf{H}_0 : \mathbf{C}\Sigma\mathbf{C}' = \lambda\mathbf{I}$, pode ser verificada da seguinte forma

$$X^2 = -\gamma \ln(\mathbf{W}),$$

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

Alguns autores como Crowder e Hand (1990) alertam que, devido esse teste envolver variâncias e covariâncias, é um teste bastante sensível à falta de normalidade dos dados.

Vonesh e Chinchilli (1997) também mostram, que o teste de esfericidade não é muito poderoso para amostras pequenas e não é robusto quando há violação de normalidade. Os autores ainda advertem que a análise de medidas repetidas torna-se mais complexa quando houver mais de um fator em estudo com medidas repetidas, por exemplo, quando se têm dois fatores intra-indivíduos, como tempo, dias e a interação tempo $\times$ dias, há necessidade, de construir um teste de esfericidade, um teste F e correções para os números de graus de liberdade para cada um desses fatores intra-indivíduos.

Caso as suposições de normalidade, independência e da condição H-F para a matriz de covariâncias não forem satisfeitas, uma alternativa seria a análise multivariada. Outra possibilidade, segundo Xavier (2000), seria utilizar a análise univariada no esquema parcela subdividida no tempo, realizando o ajuste do número de graus de liberdade do teste F para o fator intra-indivíduo, conforme já mostrado.

3.3.6 Correções do Número de Graus de Liberdade

A correção do número de graus de liberdade (G.L.) deve ser feita apenas em estatísticas que envolvem comparações “dentro” ou seja, intra-indivíduos. Portanto, a estatística F_2 não necessita de correção de seu número de grau de liberdade, qualquer que seja a estrutura de Σ . Está associada apenas a comparações “entre” ou seja inter-indivíduos; F_2

sempre terá distribuição exata F central. As estatísticas F_3 e F_4 podem precisar de correções, pois envolvem comparações intra-indivíduos, que estão associadas diretamente à estrutura da matriz.

O primeiro a propor a correção para o número de graus de liberdade foi Box (1954 I, II). O objetivo era obter uma aproximação da distribuição F, nos casos em que a matriz de covariância dos erros intra-indivíduos não leva em conta a suposição de variância constante.

Alguns anos depois, Geisser e Greenhouse (1958)²; Huynh e Feldt (1976) também sugeriram ajustes para o número de graus de liberdade do teste F para o fator erro intra-indivíduos. Os autores desenvolveram respectivamente os seguintes fatores de ajuste:

I - Ajuste de GEISSER e GREENHOUSE $\hat{\varepsilon}$

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

ou ainda admitindo: $A_{q \times q}$, com $q = t-1$ contrastes ortogonais normalizados, sobre t medidas repetidas e a_{ii} é um elemento genérico, fator de correção pode ser expresso como

$$\hat{\varepsilon} = \frac{\left(\sum_{i=1}^q a_{ii}\right)^2}{(t-1) \sum_{i=1}^q \sum_{j=1}^q a_{ij}^2}.$$

II - Ajuste de HUYNH e FELDT $\tilde{\varepsilon}$

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

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, ocasiões (tempo).

Assim, através de soluções univariadas aproximadas, as estatísticas F_3 e F_4 têm distribuições aproximadas F centrais, com $(t-1)\varepsilon$ e $g(b-1)(t-1)\varepsilon$; $(g-1)(t-1)\varepsilon$ e $g(b-1)(t-1)\varepsilon$ graus de liberdade ajustados, respectivamente.

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

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

Caso a condição de esfericidade seja satisfeita, pode-se verificar que:

$$\varepsilon = \frac{[tr(\lambda \mathbf{I}_{(t-1)})]^2}{(t-1)tr(\lambda \mathbf{I}_{(t-1)})^2} = \frac{[\lambda(t-1)]^2}{(t-1)\lambda^2(t-1)} = 1,$$

e nenhuma correção nos números de graus de liberdade precisa ser feita e a distribuição é exata.

É interessante destacar que o uso dos extremos 1 ou $(t-1)^{-1}$, como valor de ε , pode levar a teste muito liberal ou muito conservativo, respectivamente. O ajuste de HUYNH e FELDT $\tilde{\varepsilon}$ pode ser considerado mais liberal do que o ajuste de GEISSER e GREENHOUSE $\hat{\varepsilon}$. Porém, o estimador $\tilde{\varepsilon}$ para ε é menos viesado do que $\hat{\varepsilon}$ quando $\varepsilon \geq 0,75$. Nota-se também que $\tilde{\varepsilon}$ pode assumir valores maiores do que um. Nesse caso, deve-se considerar $\tilde{\varepsilon} = 1$. Uma regra geral descrita por Singer e Andrade (1986) consiste em usar o $\hat{\varepsilon}$ toda vez que não se tem idéia do verdadeiro valor de ε ou que $\varepsilon < 0,75$, e usar $\tilde{\varepsilon}$ como estimador para ε só quando se acredita que $\varepsilon \geq 0,75$.

Alguns autores como Muller e Barton (1989)⁴ citados por Xavier (2000), acreditam que com a correção do número de graus de liberdade obtêm-se testes mais conservativos, que são limitados a assegurar que α esteja abaixo de um certo nível. Isso para casos em que um teste aproximado não é desejável e nos casos onde a matriz de covariâncias é diferente de tratamento para tratamento. Os autores, depois de vários estudos com simulações, chegaram à conclusão que a correção de Geisser e Greenhouse deve ser utilizada pois o teste produz aceitável controle do erro tipo I enquanto maximiza o poder, embora de acordo Huynh e Feldt (1976) a correção de Geisser e Greenhouse possui a desvantagem de superestimar o verdadeiro nível de significância.

Então, mesmo que a condição H-F não seja satisfeita, os autores recomendam fazer a análise univariada, mas para isso o teste de esfericidade deve ser significativo ao nível de probabilidade entre 0,01 e 0,05 e com o ajuste dos graus de liberdade utilizando-se a correção de Huynh e Feldt.

⁴MULLER, K. E. e BARTON, C. N. Approximate power for repeated-measures ANOVA lacking sphericity.

Journal of the American Statistical Association, v.84, n.406, p.549-555, Jun 1989.

3.3.7 Modelos Mistos

As duas metodologias descritas para análise de dados longitudinais, indicam que o modelo univariado requer certa suposição referente à estrutura da matriz de covariâncias, e quando esta não é satisfeita leva a testes aproximados. No caso do modelo multivariado, a presença da estrutura geral para a matriz de covariâncias, torna difícil quando os dados são desbalanceados. Uma técnica alternativa aos modelos uni e multivariado, quando a suposição de esfericidade da matriz de variâncias-covariâncias não é satisfeita, há observações perdidas ou delineamento desbalanceado, é a metodologia de modelos mistos.

Essa metodologia, foi inicialmente, proposta por Henderson (1949, 1975), e tornou-se conhecida nas últimas décadas. É uma extensão do modelo linear geral e tem sido utilizada, freqüentemente, na análise de informações provenientes de experimentos ou estudos observacionais. A maior flexibilidade dos modelos mistos, de acordo com Riboldi (2007) é que ele possibilita contemplar a estrutura de interdependência das observações e permite acomodar a estrutura de correlação presente em medidas repetidas tomadas na mesma unidade experimental.

O modelo misto contempla efeitos fixos e aleatórios, tornando-se ferramenta usual na análise de experimentos em blocos ao acaso, em parcela subdividida e em delineamentos hierárquicos. Engloba análise de curvas de crescimento, ou curvas polinomiais, que levam em conta a estrutura da matriz de covariâncias que melhor explica o comportamento das observações, tendo a vantagem de englobar modelos que reduzem o número de parâmetros (VON ENDE, 1993; LITTELL et al., 1998).

Os modelos de curvas de crescimento ou polinomiais segundo Kshirsagar e Smith (1995) são mais gerais do que o modelo usual de medidas repetidas. A diferença entre eles está no interesse do pesquisador. Por exemplo, em modelos de medidas repetidas, o interesse básico é detectar diferença entre os tratamentos no decorrer do tempo. Já os modelos de curvas de crescimento, o objetivo básico é estimar e prever os efeitos de tratamentos em algum tempo.

Uma vantagem de se trabalhar com modelos mistos que envolvem curvas de crescimento ou polinomiais é a possibilidade de poder optar por uma estrutura de covariâncias que represente melhor os dados.

O princípio básico das curvas de crescimento é que existe uma relação funcional entre os efeitos de tratamentos e o tempo, e esta função pode ser modelada e aproximada por um polinômio, cujos coeficientes, variâncias e covariâncias, podem ser estimados. No processo de modelagem é permitida a utilização de várias estruturas de covariâncias.

O modelo misto é especificado pela seguinte fórmula matricial:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\beta_i + \mathbf{Z}_{ij}\mathbf{u}_{ij} + \varepsilon_{ij}$$

para $i = 1, 2, \dots, g$ e $j = 1, 2, \dots, n_i$, em que $\mathbf{y}_{ij}(p \times 1)$ é o perfil de respostas do indivíduo (ij) , $\mathbf{X}_{ij}(p \times r)$ é uma matriz de posto $r < p$, conhecida e de especificação, associada ao vetor $\beta_i(r \times 1)$ de parâmetros fixos desconhecidos, $\mathbf{Z}_{ij}(p \times q)$ é uma matriz conhecida e de especificação, de posto coluna completo, associada ao vetor de parâmetros de efeitos aleatórios desconhecidos $\mathbf{u}_{ij}(q \times 1)$ e $\varepsilon_{ij}(p \times 1)$ é um vetor de erros aleatórios.

Supõe-se que $\mathbf{u}_{ij}$ e ε_{ij} são independentes, que $\varepsilon_{ij} \sim N(0, I\sigma^2 = \mathbf{R}_{ij})$ e que $\mathbf{u}_{ij} \sim N(0, A\sigma_a^2 = \mathbf{G})$, para $i = 1, 2, \dots, g$ e $j = 1, 2, \dots, n_i$, cujas médias e matrizes de variâncias e covariâncias são dadas por:

$$E \begin{bmatrix} \mathbf{u} \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ e } Var \begin{bmatrix} \mathbf{u} \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix}$$

Os modelos mistos (2) em dois estágios, apresentados por Laird e Ware (1982); Diggle et al. (1998), consideram os efeitos fixos no primeiro estágio para obtenção da curva polinomial média, e no segundo estágio permitem diferentes curvas para cada indivíduo.

Desta forma, tem-se que:

$$\mathbf{V}_{ij} = Var(\mathbf{Y}_{ij}) = Var(\mathbf{X}_{ij}\beta_i + \mathbf{Z}_{ij}\mathbf{u}_{ij} + \varepsilon_{ij}) = \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}' + \mathbf{R}_{ij}$$

Assume-se ainda que $\mathbf{V}_{ij}$ é não singular, e

$$E(\mathbf{Y}_{ij}) = E(\mathbf{X}_{ij}\beta_i + \mathbf{Z}_{ij}\mathbf{u}_{ij} + \varepsilon_{ij}) = \mathbf{X}_{ij}\beta_i.$$

Assim,

$$\mathbf{Y}_{ij} \sim N(\mathbf{X}_{ij}\beta_i; \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}' + \mathbf{R}_{ij}).$$

É interessante ressaltar que os elementos do vetor u_{ij} são adicionados ao modelo com o objetivo de promover um melhor ajuste. É um vetor de parâmetros individuais que representa as diferenças entre a curva do tratamento e a curva de cada indivíduo.

As estimativas dos parâmetros do modelo, $\hat{\beta}$ e a predição dos efeitos aleatórios $\hat{\mathbf{u}}$ podem ser obtidas pela solução do sistema de equações do modelo misto de acordo com Henderson (1984).

3.4 Estimação em Modelos Mistos

Para estimar parâmetros e prever efeitos aleatórios, a estrutura especial da distribuição normal para $\mathbf{u}$ e ε permite dividir o desenvolvimento teórico em duas etapas. Na primeira, consideram-se $\mathbf{G}$ e $\mathbf{R}$ conhecidas, obtendo-se solução para β e $\mathbf{u}$. Depois estimam-se os parâmetros não conhecidos em $\mathbf{G}$ e $\mathbf{R}$.

3.4.1 Estimação de β e predição de $\mathbf{u}$

A obtenção das equações normais para modelos mistos pode ser feita pela minimização da soma de quadrados dos resíduos ou pela maximização da função densidade de probabilidade conjunta de $\mathbf{Y}$ e $\mathbf{U}$.

A função densidade de probabilidade de $\mathbf{y}$ é dada por:

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

A função densidade de probabilidade conjunta de $\mathbf{Y}$ e $\mathbf{U}$ pode ser escrita como o produto entre a função densidade condicional de $\mathbf{Y}$, dado $\mathbf{U}$, e a função densidade de probabilidade de $\mathbf{U}$, conforme segue

$$f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}) \cdot f(\mathbf{u})$$

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

sendo $|\mathbf{R}|$ e $|\mathbf{G}|$ os determinantes das matrizes de covariâncias.

Para proceder à maximização de $f(\mathbf{y}, \mathbf{u})$, pode-se usar a transformação por logaritmo. Assim o logaritmo da função de verossimilhança é

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

Derivando-se $l(\beta, \mathbf{u}|\mathbf{y})$ em relação a β e $\mathbf{u}$, e tornando-se tais derivadas identicamente nulas, obtêm-se:

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

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

e então

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

Essas são as equações de modelos mistos (EMM) de Henderson, que permitem obter soluções para os efeitos fixos $\hat{\beta}$ e predições para os efeitos aleatórios $\hat{u}$. As EMM são equações normais estendidas ou também equações de mínimos quadrados generalizados.

A estimação é mais complexa no modelo misto do que no modelo linear geral, pois além de β , agora o modelo tem parâmetros desconhecidos em $\mathbf{u}$, $\mathbf{G}$ e $\mathbf{R}$. Neste caso, o método de estimação dos mínimos quadrados não é o melhor, sendo o método de mínimos quadrados generalizado (GLS) mais apropriado, minimizando $(y - X\beta)'V^{-1}(y - X\beta)$.

As soluções são dadas por

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (2)$$

que podem ser escritas como

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

que é a mesma solução de mínimos quadrados generalizados, e

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

O estimador de β e o preditor de $\mathbf{u}$, são respectivamente, o estimador de mínimos quadrados generalizados (GLSE) e o melhor preditor linear imparcial (BLUP).

3.4.2 Estimando os parâmetros de covariância em G e R

Na maioria dos casos a matriz de variâncias e covariâncias V é desconhecida, assim como, G e R são desconhecidas, sendo necessário obter suas estimativas. As estimativas de G e R podem ser obtidas através dos métodos da Máxima Verossimilhança (ML) e Máxima Verossimilhança Restrita (REML), aplicados a $(y - X\beta)'V^{-1}(y - X\beta)$. Para obter as estimativas de máxima verossimilhança são usados algoritmos iterativos tais como o de Newton-Rapson, Score de Fisher e o algoritmo EM.

- i. Método da Máxima Verossimilhança (ML), cujo log da função de verossimilhança é o seguinte:

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

- ii. Método da Máxima Verossimilhança Restrita (REML), cujo log da função de verossimilhança é expressa por:

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

em que p é a característica da matriz X .

Nos casos onde V é desconhecida, consequentemente, as matrizes G e R também são, assumindo-se alguma perda de eficiência, as estimativas $\hat{G}$ e $\hat{R}$ substituem os seus respectivos parâmetros na expressão (2) então, a estimação do parâmetro β e a predição de $\mathbf{u}$ são obtidas através de

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

em que $\hat{C}$ é a matriz de covariância estimada, que pode ser expressa como

$$\hat{C} = \begin{bmatrix} \hat{C}_{11} & \hat{C}'_{21} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix}$$

com

$$\begin{aligned} \hat{C}_{11} &= (X'\hat{V}^{-1}X)^{-} \\ \hat{C}_{21} &= -\hat{G}Z'\hat{V}^{-1}X\hat{C}_{11} \\ \hat{C}_{22} &= (Z'\hat{R}^{-1}Z + \hat{G}^{-1})^{-1} - \hat{C}_{21}X'\hat{V}^{-1}\hat{G}. \end{aligned}$$

Nota-se que $\hat{C}_{11}$ é uma fórmula já conhecida, para estimar a matriz de variâncias e covariâncias do estimador de mínimos quadrados generalizado de β , isto é, $\hat{V}(\hat{\beta}) = \hat{C}_{11} = (X'\hat{V}^{-1}X)^{-}$.

Quando os parâmetros G e R são substituídos pelas suas respectivas matrizes de covariâncias estimadas $\hat{G}$ e $\hat{R}$, o termo empírico ou estimado é acrescentado a BLUE e BLUP, obtendo-se EBLUE (*empirical best linear unbiased estimator*) e EBLUP (*empirical best linear unbiased predictor*), de acordo com Littell et. al. (1996).

3.5 Estruturas da Matriz de Covariâncias

Mostrou-se, em tópicos anteriores, que as três técnicas abordadas para análise de dados longitudinais apresentam estruturas da matriz de covariância diferentes. No caso Multivariado, a matriz de covariância é não estruturada, assume uma forma mais geral, permitindo diferentes variâncias e covariâncias. Nos modelos univariados é exigido que a estrutura da matriz de covariância satisfaça a condição H-F, conforme já comentado. Quando se trata de modelos mistos, é possível modelar a estrutura da matriz de covariância da forma que melhor represente os dados respeitando sua natureza, sejam eles independentes, dependentes, correlacionados ou que apresente qualquer outro tipo de relação que não se consegue explicar utilizando estrutura da matriz de covariância usual.

Contudo, várias são as estruturas da matriz de covariâncias comumente utilizadas. Além da matriz H-F, já apresentada, a Tabela 3 ilustra algumas estruturas que estão implementas no SAS.

Tabela 3 - Estrutura da Matriz de Covariâncias, Características

Estruturas	Características
Simétrica Composta (CS)	$\begin{bmatrix} (\sigma^2 + \sigma_1^2) & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & (\sigma^2 + \sigma_1^2) & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & (\sigma^2 + \sigma_1^2) & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & (\sigma^2 + \sigma_1^2) \end{bmatrix}$
Não Estruturada (UN)	$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}$
Auto-Regressiva 1ª Ordem (AR (1))	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$
Componentes de Variância (VC)	$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$
Toepliz (TOEP)	$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$
Simétrica Composta Heterogênea (CSH)	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$
Auto-Regressiva Heterogênea (ARH (1))	$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$

3.6 Seleção do Modelo e Teste Estatístico

A seleção do modelo é um tópico complexo devido ao grande número de estruturas de covariâncias. Após construir alguns modelos plausíveis de representação das observações e escolher as estruturas de covariância para os efeitos aleatórios, alguns métodos de seleção podem ser utilizados para selecionar o modelo que melhor represente os dados do experimento. Dentre os métodos utilizados na seleção do modelo destacam-se

i.) Comparações de Modelos via Teste Assintótico da Razão de Verossimilhança - TARV.

O princípio desse teste consiste em comparar dois modelos estimados por máxima verossimilhança, em que um deles é uma versão restrita do outro, ou seja, um modelo tem r parâmetros adicionais. O teste irá verificar se esses parâmetros adicionais melhoram significativamente o modelo. Definindo-se ℓ_1 , o valor de $\ell = (-2 \ln$ da função de verossimilhança) para o modelo com o menor número de parâmetros e ℓ_2 , para o modelo com maior número de parâmetros, ou seja, para o modelo com r parâmetros extras, a hipótese a ser testada é a de que os dois modelos são equivalentes (os parâmetros extras não diferem de zero). A diferença entre os valores ℓ_1 , e ℓ_2 , é assintoticamente distribuída como uma Qui-Quadrado com r graus de liberdade (MOOD et al., 1974). A estatística resultante dessa diferença tem uma distribuição amostral, que segue uma distribuição qui-quadrado com r graus de liberdade dados pela diferença entre o número de parâmetros inseridos nas matrizes de covariância dos dois modelos. Logo,

$$\ell_1 - \ell_2 = \chi_r^2$$

Embora esse teste seja considerado bastante eficaz, possui a desvantagem que só pode ser usado para comparar dois modelos de cada vez, sendo que um desses modelos é sempre um caso especial do outro (GUIMARÃES, 1994; XAVIER, 2000).

ii.) Comparações de Modelos via Critério de Informação

É uma alternativa para o teste de verossimilhança. Esse procedimento de seleção consiste em minimizar os critérios de informação de máxima verossimilhança restrita. Assim, será selecionado aquele que tiver o menor valor para tais critérios. Esses critérios estão fundamentados na teoria da decisão e penaliza os modelos com número grande de parâmetros para evitar excesso de parametrização. Os mais comuns destes são os Critérios de Akaike

(AIC) e Schwarz (BIC). Os dois critérios podem ser expressos pelas seguintes fórmulas

$$AIC_R = -2\ell + 2p \quad (3)$$

$$BIC_R = -2\ell + p \log(n - posto(\mathbf{X})) \quad (4)$$

em que ℓ é o ln da função de verossimilhança, p é o número de parâmetros da matriz de covariância e n é o número de observações.

No proc MIXED do SAS, a partir da versão 8, os critérios de informações foram alterados para as fórmulas (3) e (4), do AIC_R e BIC_R , respectivamente. E os menores valores dessas estatísticas passaram a indicar os melhores modelos.

Para selecionar o modelo adequado é necessário calcular o AIC e BIC para todos os modelos considerados. É importante salientar que o critério BIC penaliza os modelos com maior número de parâmetros e que nem sempre os critérios (AIC_R e BIC_R) concordam em indicar o modelo melhor. Alguns autores afirmam que a performance do AIC é melhor que do BIC para identificar o melhor modelo.

Sabe-se que, para o modelo misto a inferência é realizada sobre os parâmetros dos efeitos fixos e aleatórios. Os testes estatísticos usados para avaliar os modelos incluem o teste Z de **Wald**, que serve para avaliar a significância dos efeitos fixos e os testes da razão de verossimilhanças, que são usados para avaliar tanto os efeitos fixos quanto os aleatórios em modelos encaixados, por intermédio da comparação de modelos mais simples com os mais gerais.

A estatística de **Wald** é calculada dividindo a estimativa do parâmetro pelo seu erro padrão assintótico. Os erros padrões assintóticos são calculados da inversa da matriz de segundas derivadas do logaritmo da verossimilhança em relação a cada um dos parâmetros de covariância.

Já o teste da razão de verossimilhanças, consiste em comparar dois modelos que são aninhados (encaixados ou modelo completo versus modelo reduzido) em relação aos parâmetros de covariância. Nesse caso, o modelo reduzido é constituído do modelo completo, impondo uma ou mais restrições, o que resulta em modelos aninhados ou encaixados.

Então, quando os modelos são encaixados e a estimação dos parâmetros é feita através do método MV (máxima verossimilhança), o teste da razão de verossimilhança é

usado para testar a hipótese nula, de que o modelo mais simples é o adequado. E quando os modelos não são encaixados, pode-se utilizar o critério de informação de akaike.

Pode-se ainda considerar combinações lineares estimáveis da seguinte forma

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

tem-se que $L\beta$ caracteriza uma função estimável para efeitos fixos, e neste caso as partes correspondentes aos parâmetros dos efeitos aleatórios u em L assumem valores zeros.

As hipóteses consideradas, para aplicação das inferências estatísticas pelo teste de hipótese, é a seguinte:

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

e quando L é constituída por uma única linha o teste de hipótese a ser aplicado para testar a hipótese dada é o teste t de Student, construído da seguinte forma:

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

Para o caso de L ser constituída por mais de uma linha, ou seja, quando o posto (L) for maior que 1, o teste de hipótese a ser aplicado para testar a hipótese dada é o teste F , construído conforme segue

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

Analogamente à t , F , em geral, tem uma distribuição F aproximada com número de graus de liberdade no numerador igual ao $\text{posto}(L)$. As estatísticas t e F permitem fazer inferências sobre os efeitos fixos, estimados para o modelo de variância e covariância selecionado.

4 MATERIAL E MÉTODOS

Para alcançar o objetivo proposto nesse trabalho, utilizou-se dados de dois experimentos de cana-de-açúcar da Rede Interuniversitária de Desenvolvimento do Setor Sucroalcooleiro-RIDESA. Foram estudados o comportamento de 48 genótipos RB (República do Brasil) da série 93, 94, 95 e 96, em dois experimentos, o Experimento 1 que foi instalado na Usina Itamarati, em São Paulo e o Experimento 2 instalado na Usina Porto Rico, em Alagoas. Esses ensaios foram conduzidos pelos Programas de Melhoramento Genético da Cana-de-açúcar, PMGCA-UFSCar e PMGCA-UFAL, das Universidade Federal de São Carlos e Universidade Federal de Alagoas, respectivamente.

O Experimento 1 foi instalado em 05/05/2003, em uma das unidades da Usina Itamarati, possui 28 genótipos, no delineamento experimental de blocos ao acaso, com três repetições, parcelas de $60m^2$ e espaçamentos de $1,50m$ entre linhas. Esse ensaio foi conduzido no período de maio de 2003 a agosto de 2006, obtendo-se assim, informações de um período de 3 anos, ou seja, avaliou-se três colheitas, com dados de cana-planta, soca e ressoca. O Experimento 2, foi instalado em 15/09/2001, na Fazenda Santa Tereza, unidade da Usina Porto Rico, com 20 genótipos, no mesmo delineamento do experimento anterior, porém com quatro repetições, parcelas de $50,4m^2$ e espaçamento de $1,40m$ entre linhas. Conduzido no período de setembro de 2001 a outubro de 2006 e foram obtidas informações de cinco colheitas, avaliando assim, a cana-planta, soca e três ressocas, em cinco anos seguidos.

Nesse estudo, foi analisada a variável TCH - Tonelada de Cana por Hectare. Por meio dessa informação foi estudado o perfil médio de resposta dos 48 genótipos de cana-de-açúcar. E através da metodologia de dados longitudinais (análise de medidas repetidas), analisou-se os dados utilizando-se os modelos: univariado, multivariado e misto, já discutidos nesse contexto. Para análise foram utilizados dois procedimentos do SAS, o proc GLM e o proc MIXED. A princípio sabe-se que, tanto no modelo univariado como no modelo multivariado, o proc GLM utiliza uma estrutura tradicional na obtenção dos resultados. Porém o proc GLM requer que os dados sejam balanceados, devido utilizar o método dos momentos na estimação dos efeitos. Para isso, é necessário que as observações a respeito de cada indivíduo sejam completas, isto é, sem nenhum valor perdido. Essa característica torna o proc GLM limitado, portanto, em situações para os quais ocorrem desbalanceamento dos dados, o proc

MIXED é mais indicando, especialmente quando se trata de modelos mistos.

O proc GLM antes de proceder a análise identificará os genótipos que apresentarem informações perdidas, selecionando-se um modelo de médias, em termos de efeitos fixos da parcela e da subparcela, entre e intra-indivíduos, respectivamente. Nesse estudo o efeito da parcela (entre-indivíduos) são os genótipos, o efeito da subparcela (dentro-indivíduo) são os anos e a interação genótipo $\times$ ano. É necessário especificar esses efeitos separadamente quando se adota esse procedimento.

Para análise de dados com medidas repetidas pelo proc GLM é necessário que seja indicada uma transformação (tipo de contraste) já que um conjunto de contrastes pode ser usado para analisar tendências sobre o fator de medidas repetidas, e realizar comparações entre os níveis desse mesmo fator.

Nesse procedimento, os dados originais de cada genótipo são transformados em um novo conjunto de dados, obtidos através de um conjunto de $t - 1$ contrastes, em que t é o número de medidas repetidas. Essas transformações (contrastos) são utilizadas com intuito de amenizar a influência de algumas estruturas de covariâncias, na análise univariada de medidas repetidas, que podem invalidar os resultados dos testes.

Para efeitos fixos, no caso de parcela, o proc GLM apresenta um teste padrão. E para os efeitos de subparcela existem testes univariados, que são válidos quando a condição H-F é satisfeita. Essa suposição será verificada pelo teste de esfericidade, que está implementado no proc GLM do SAS. Então, quando o teste de esfericidade for significativo, pode-se proceder a correção dos graus de liberdade, para efeito de ano e genótipo $\times$ ano.

Outra alternativa é a utilização de testes multivariados como: Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Roy. Esses testes são todos baseados numa matriz de covariância sem estrutura, e assim realizar a análise multivariada.

Uma terceira alternativa é a análise de modelo misto pelo proc MIXED, sendo que, para esse estudo, será considerado o efeito de blocos como aleatórios e os demais efeitos como fixos. O proc MIXED do SAS utiliza os métodos da máxima verossimilhança e máxima verossimilhança restrita na estimação dos parâmetros, permitindo a inclusão de genótipos que tenham alguma informação perdida. Esse procedimento dispensa identificação inicial dos efeitos entre e dentro de indivíduos, determina apenas, o modelo de médias, sem a necessidade

de trabalhar com uma transformação para os dados. Nesse processo é preciso especificar a estrutura da matriz de covariâncias que melhor represente os dados. Os testes para os efeitos fixos são baseados nessa estrutura da matriz de covariâncias, por isso, caso haja seleção de um modelo de matriz não adequado, isso implicará em resultados não seguros.

O processo de seleção dos modelos contendo as várias estruturas de covariâncias pode ser realizado por vários critérios já implementados no SAS, dentre eles pode-se utilizar o AIC (critério de informação de Akaike), BIC (Critério Bayesiano de Schwarz), entre outros, ou então, construindo-se um teste de razão de verossimilhança, entre estruturas de covariâncias duas a duas, mas com a condição de que essas estruturas sejam casos especiais umas das outras (dentro do par).

Neste trabalho, foram abordados os métodos univariado, multivariado e misto, destacando algumas estruturas da matriz de covariâncias e um tópico sobre a seleção dos modelos comparados via Critério de Informação de Akaike (AIC) e via BIC (Critério Bayesiano de Schwarz).

5 RESULTADOS E DISCUSSÃO

Para o Experimento 1 foram obtidos os seguintes resultados:

5.1 Análise Univariada

As Tabelas 4 e 5 apresentam os resultados da análise univariada e teste de esfericidade, via proc GLM, no esquema de parcelas subdivididas.

Observa-se que existem diferenças significativas dos fatores Genótipo, Anos e da interação Genótipo $\times$ Anos. Isso significa que em relação a TCH, os genótipos são diferentes, ou seja, os perfis são não coincidentes, não horizontais e também não paralelos, respectivamente. Mas, para assegurar a tomada de decisão é necessário verificar se a condição de esfericidade foi satisfeita através da estatística de Muchly (1940), mostrada em 3.3.5.

Tabela 4 - Resultado da Análise Univariada Usando o proc GLM

Variável Dependente: TCH					
Causas de Variação	G.L.	S.Q.	Q.M.	F	Pr > F
Modelo	139	64400,96862	463,31632	15,02	< 0,0001
Resíduo	112	3455,41193	30,85189		
Total Corrigido	251	67856,38056			
Bloco	2	154,74636	77,37318	2,51	0,0860
Genótipo	27	36665,36456	1357,97647	44,02	< 0,0001
Genótipo × Bloco	54	2890,79150	53,53318	1,74	0,0074
Anos	2	13399,85203	6699,92601	217,16	< 0,0001
Genótipo × Anos	54	11290,21417	209,07804	6,78	< 0,0001
$R^2 = 0,9491$ C.V. = 4,58 % Média de TCH = 121,279					

Tabela 5 - Teste de Esfericidade

Hipótese	G.L.	Critério de Aproximação	Prob. > $\chi^2_{(0,05;2)}$
		Mauchly(W) Qui-Quadrado	
$H_0: C\Sigma C' = \lambda I$	2	0,7870	12,6927
			0,0018

Pode-se observar na Tabela 5, que a condição de esfericidade foi violada com um nível de significância de 0,0018. Isso significa que a matriz de covariâncias não pode ser considerada do tipo Huynh-Feldt. Uma vez que suposição de esfericidade não foi atendida, conforme os autores: Box (1954 I, II); Geisser e Greenhouse (1958); Huynh e Feldt (1976), é necessário realizar as correções para os números de graus de liberdade dos fatores intra-indivíduos, ou seja, para o fator Anos e a interação Genótipo × Anos.

5.2 Análise Univariada e Correção dos Números de Graus de Liberdade

Na Tabela 6 encontram-se os resultados da análise univariada com a correção dos números de graus de liberdade.

A hipótese de perfis coincidentes é verificada através do teste para o fator entre indivíduos, ou seja, o fator Genótipo, que é rejeitado, indicando que as distâncias médias

Tabela 6 - Análise Univariada e Correção dos Números de Graus de Liberdade para os Efeitos Intra-indivíduo

TESTE PARA OS FATORES ENTRE INDIVÍDUOS							
C.V.	G.L.	S.Q.	Q.M.	F	Pr > F		
Bloco	2	154,74636	77,37318	1,45	0,2446		
Genótipo	27	36665,36456	1357,97647	25,37	<0,0001		
Resíduo	54	2890,7915	53,53318				
TESTE PARA OS FATORES INTRA INDIVÍDUOS							
C.V.	G.L.	S.Q.	Q.M.	F	Pr > F	G-G	H-F
Anos	2	13399,852	6699,926	215,32	<0,0001	<0,0001	<0,0001
Anos × Bl.	4	94,900	23,723	0,76	0,5519	0,5294	0,5519
Gen. × Anos	54	11290,214	209,078	6,72	<0,0001	<0,0001	<0,0001
Res. (Ano)	108	3360,511	31,116				
Greenhouse-Geisser $\hat{\varepsilon} = 0,8244$							
Huynh-Feldt $\tilde{\varepsilon} = 1,3037$							

dos genótipos são diferentes ($p < 0,0001$). Já hipótese de perfis paralelos, pode ser verificada através do teste para a interação Genótipo × Anos, que também foi rejeitada, o mesmo ocorreu com a hipótese de perfis horizontais, que é verificada pelo fator Anos.

Observa-se também que foram fornecidas as correções dos números de graus de liberdade dos testes F para os fatores intra-indivíduos. Como a condição de esfericidade foi rejeitada, para tomada de decisão em relação às hipóteses, os níveis mínimos de significância em negrito, das estatísticas Greenhouse-Geisser ($\hat{\varepsilon} = 0,8244$) e Huynh-Feldt ($\tilde{\varepsilon} = 1,3037$) foram considerados. Nota-se que, a correção de Huynh-Feldt foi superior a um, então, conforme sugerido por Box (1954 I, II); Singer e Andrade (1986), destacados em 3.3.6, considerou-se $\tilde{\varepsilon} = 1$, que corresponde a condição de esfericidade. Um detalhe a ser ressaltado é que, mesmo com as correções para os números de graus de liberdade, a significância dos testes não foi alterada.

5.3 Análise Multivariada

Os resultados da Análise Multivariada via proc GLM, fornecem as matrizes E, de soma de quadrados e produtos cruzados do resíduo (SSPC) e R, das correlações parciais dos resíduos com os respectivos níveis descritivos associados, conforme verificam-se

$$E = \begin{bmatrix} & \textit{ano1} & \textit{ano2} & \textit{ano3} \\ \textit{ano1} & 3038,093 & 157,669 & 418,414 \\ \textit{ano2} & 157,669 & 798,549 & 634,452 \\ \textit{ano3} & 418,414 & 634,452 & 2414,660 \end{bmatrix}$$

$$R = \begin{bmatrix} & \textit{ano1} & \textit{ano2} & \textit{ano3} \\ \textit{ano1} & 1,0000 & 0,1012 & 0,1545 \\ & 0,000 & 0,462 & 0,260 \\ \textit{ano2} & 0,1012 & 1,0000 & 0,4569 \\ & 0,462 & 0,000 & 0,000 \\ \textit{ano3} & 0,1545 & 0,4569 & 1,0000 \\ & 0,260 & 0,000 & 0,000 \end{bmatrix}$$

A Matriz R mostra as correlações entre as observações tomadas nas três cortes sucessivos da cana-de-açúcar durante três anos, e abaixo de cada correlação têm-se seus respectivos níveis de significância em *italico*. Verifica-se que as observações tomadas entre os anos 1 e 2, 1 e 3, 2 e 3, tiveram correlações iguais a 0,1012, 0,1545 e 0,4569, respectivamente. Esse resultado contraria a hipótese apresentada por Littell et al. (1998); Xavier (2000); Costa (2003) e Rosário (2003), de que as resposta de tempo mais próximo são, em geral, mais fortemente correlacionada.

Por meio dos testes multivariados, foram testados os fatores Anos e Genótipo $\times$ Anos. Os resultados da Tabela 7 mostram que, na primeira parte foram usadas as matrizes H, da soma de quadrados e produtos cruzados para o fator Anos e E, da soma de quadrados e produtos cruzados do resíduo. Além os parâmetros, $S = 1$, $m_1 = 0$ e $m_2 = 25,5$, que de acordo com Timm (1975); Morrison (1990) proporcionam a distribuição exata das estatísticas.

Tabela 7 - Resultado da Análise Multivariada Usando o proc GLM

H = Matriz da soma de quadrados e produtos cruzados para o fator Anos					
E = Matriz da soma de quadrados e produtos cruzados do resíduo					
$S = 1 \ m_1 = 0 \ m_2 = 25,5$					
Estatística	Valor	F	Num. G.L. ¹	Den. G.L. ²	Pr > F
Lambda de Wilks	0,0958	250,01	2	53	< 0,0001
Traço de Pillai	0,9041	250,01	2	53	< 0,0001
Traço de Hotelling-Lawley	9,4344	250,01	2	53	< 0,0001
Raiz de Roy	9,4344	250,01	2	53	< 0,0001
H = Matriz da soma de quadrados e produtos cruzados para o fator Genótipo × Anos					
E = Matriz da soma de quadrados e produtos cruzados do resíduo					
$S=2 \ m_1 = 12 \ m_2 = 25,5$					
Lambda de Wilks	0,0467	7,12	54	106	< 0,0001
Traço de Pillai	1,5586	7,06	54	108	< 0,0001
Traço de Hotelling-Lawley	7,4479	7,19	54	93,967	< 0,0001
Raiz de Roy	4,6791	9,36	27	54	< 0,0001

1: Grau de liberdade do numerador; 2: Grau de liberdade do denominador.

Na segunda parte da análise têm-se as matrizes H, da soma de quadrados e produtos cruzados para o fator Genótipo × Anos e E, da soma de quadrados e produtos cruzados do resíduo. E também os parâmetros $S = 2$, $m_1 = 12$ e $m_2 = 25,5$.

Nota-se que, para o fator Ano, os quatro testes, Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz de Roy, foram altamente significativos. Isso significa que a hipótese de perfis horizontais, é rejeitada. Observa-se também que a interação Genótipo × Anos, que testa a hipótese de paralelismo, também é rejeitada, e conforme pode ser verificado nas Tabelas 4 e 6, esses resultados são os mesmos obtidos pela análise univariada.

5.4 Análise Via Modelo Misto

Para selecionar a estrutura da Matriz de Covariâncias foram testados oito modelos descritos na Tabela 8.

Tabela 8 - Estrutura da Matriz de Covariâncias, Parâmetros Estimados e Critérios de Informação de Akaike (AIC) e de Schwarz (BIC)

Estruturas da Matriz de Covariâncias	Parâmetros	AIC (Akaike)	BIC (Schwarz)
Simétrica Composta: CS	2	1174,1	1179,0
Não Estruturada: UN	6	1151,6	1166,2
Auto-Regressiva 1ª Ordem: AR (1)	2	1174,4	1179,2
Componentes de Variância: VC	1	1178,0	1180,4
Toepliz: TOEP	3	1175,5	1182,8
Huynh-Feldt: HF	4	1159,2	1168,9
Simétrica Composta Heterogênea: CSH	5	1152,7	1162,4
Auto-Regressiva Heterogênea: ARH (1)	4	1153,0	1162,7

Conforme apresentado em 3.6, para seleção do modelo da estrutura da matriz de covariância, consideraram os critérios de AIC e BIC. Observa-se que de acordo com o critério de AIC a estrutura selecionada seria a UN, pois apresenta menor AIC. Porém o modelo UN apresenta o maior número de parâmetros a ser estimado, o que é inconveniente. Nesse caso, a estrutura CSH pode ser considerada uma solução mais plausível, pois apresenta menor BIC e o segundo menor AIC, e com o número de parâmetro a ser estimado é igual a cinco.

A Tabela 9 apresenta os testes para os efeitos fixos do modelo quando a estrutura da matriz de covariâncias selecionada é do tipo CSH.

Tabela 9 - Teste para os Efeitos Fixos do Modelo com a Estrutura da Matriz de Covariâncias do Tipo Simétrica Composta Heterogênea

Causas de Variação	Num. G.L.	Den. G.L.	F	Pr > F
Bloco	2	54	0,98	0,3824
Genótipo	27	54	24,30	< 0,0001
Anos	2	112	216,09	< 0,0001
Genótipo $\times$ Anos	54	112	6,87	< 0,0001

Nesse caso, observa-se que os efeitos Genótipos, Anos e interação Genótipos $\times$ Anos foram significativos, portanto, devem permanecer no modelo.

De acordo com o exposto, o perfil médio de resposta, em função da produtividade (TCH) dos 28 genótipos estudados em um período de três anos, pode ser representado graficamente conforme a Figura 1.

Figura 1 - Perfis médios de TCH dos 28 genótipos durante três anos

Nota-se na Figura 1 os perfis médios da produtividade dos genótipos nas diferentes condições de avaliação (anos), que por meio de análise visual é possível verificar as

rejeições das hipóteses dos perfis serem horizontais, paralelos e coincidentes, ocorridas nos três modelos de análise de dados longitudinais empregados.

As mesmas análises foram realizadas para no Experimento 2, conforme pode ser verificado a seguir.

5.5 Análise Univariada

As Tabelas 10 e 11 apresentam os resultados da análise univariada e teste de esfericidade respectivamente, via proc GLM, no esquema de parcelas subdivididas.

Tabela 10 - Resultado da Análise Univariada Usando o proc GLM

Variável Dependente: TCH					
Causas de Variação	G.L.	S.Q.	Q.M.	F	Pr > F
Modelo	159	134414,7862	845,3760	18,26	< 0,0001
Resíduo	240	11112,6483	46,3027		
Total Corrigido	399	145527,4345			
Bloco	3	1110,95433	370,31811	8,00	< 0,0001
Genótipo	19	28266,39416	1487,70496	32,13	< 0,0001
Genótipo × Bloco	57	3152,30913	55,30367	1,19	0,1817
Anos	4	81197,16950	20299,29238	438,40	< 0,0001
Genótipo × Anos	76	20687,95907	272,20999	5,88	< 0,0001
$R^2 = 0,92$ C.V. = 7,46% Média de TCH = 91,186					

Observa-se que, em relação aos hipóteses de interesse, pode-se afirmar que existem diferenças significativas dos fatores Genótipo, Anos e da interação Genótipo × Anos. Logo, semelhantemente ao experimento anterior, os perfis são não coincidentes, não horizontais e não paralelos, respectivamente. E para assegurar a tomada de decisão é necessário verificar se a condição de esfericidade foi satisfeita.

Nota-se que, a hipótese de perfis coincidentes é rejeitada. E as hipóteses de perfis paralelos, verificada através do teste para a interação Genótipo $\times$ Anos e de perfis horizontais, obtida com o teste para o fator Anos, também foram.

Verifica-se ainda, que foram fornecidas as correções dos números de graus de liberdade dos testes F para os fatores intra-indivíduos, e que os níveis mínimos de significância em negrito foram considerados na tomada de decisão, mas semelhantemente ao resultado do experimento anterior, a correção de Huynh-Feldt foi maior que um, portanto, considerou-se $\tilde{\varepsilon} = 1$, e a significância dos testes não foi alterada. Porém, apesar da conclusão ser comum aos dois experimentos estudados, esse fato não descarta a necessidade de correção dos números de graus de liberdade em outros experimentos.

5.7 Análise Multivariada

Na Análise Multivariada via proc GLM, foram obtidas as matrizes E, de soma de quadrados e produtos cruzados do resíduo (SSPC) e R, das correlações parciais dos resíduos com os respectivos níveis descritivos associados, conforme verifica-se

$$E = \begin{bmatrix} & \textit{ano1} & \textit{ano2} & \textit{ano3} & \textit{ano4} & \textit{ano5} \\ \textit{ano1} & 1165,673 & 287,988 & 60,399 & -214,360 & -54,133 \\ \textit{ano2} & 287,988 & 1877,586 & 450,739 & 376,871 & -87,589 \\ \textit{ano3} & 60,399 & 450,739 & 3701,341 & -869,245 & 1,962 \\ \textit{ano4} & -214,359 & 376,871 & -869,245 & 3511,666 & 1549,505 \\ \textit{ano5} & -54,132 & -87,589 & 1,962 & 1549,505 & 2501,002 \end{bmatrix}$$

$$R = \begin{bmatrix} & \text{ano1} & \text{ano2} & \text{ano3} & \text{ano4} & \text{ano5} \\ \text{ano1} & 1,0000 & 0,1947 & 0,0290 & -0,1059 & -0,0317 \\ & 0,000 & 0,143 & 0,828 & 0,429 & 0,813 \\ \text{ano2} & 0,1947 & 1,0000 & 0,1710 & 0,1468 & -0,0404 \\ & 0,143 & 0,000 & 0,199 & 0,271 & 0,763 \\ \text{ano3} & 0,0290 & 0,1710 & 1,0000 & -0,2411 & 0,0006 \\ & 0,828 & 0,199 & 0,000 & 0,068 & 0,068 \\ \text{ano4} & -0,1059 & 0,1468 & -0,2411 & 1,0000 & 0,5228 \\ & 0,429 & 0,271 & 0,068 & 0,000 & 0,000 \\ \text{ano5} & -0,0317 & -0,040 & 0,0006 & 0,5228 & 1,0000 \\ & 0,813 & 0,763 & 0,996 & 0,000 & 0,000 \end{bmatrix}$$

Nota-se que, na matriz R as observações tomadas entre os anos 4 e 5 foram mais fortemente correlacionadas e significativa. Verifica-se ainda, que a hipótese de que observações tomadas em tempos mais próximos são, em geral, mais fortemente correlacionada foi contrariada.

Observando-se os testes multivariados na Tabela 13, pode-se notar que, da mesma forma que no experimento anterior, têm-se na primeira parte as matrizes H, da soma de quadrados e produtos cruzados para o fator Anos e E, da soma de quadrados e produtos cruzados do resíduo. Sendo os parâmetros, $S = 1$, $m_1 = 1$ e $m_2 = 26$. E na segunda parte usou-se as matrizes H, da soma de quadrados e produtos cruzados para o fator Genótipo $\times$ Anos e E, da soma de quadrados e produtos cruzados do resíduo. E com $S = 4$, $m_1 = 7$ e $m_2 = 26$.

Para o fator Ano, a hipótese de perfis horizontais, é rejeitada, pois as estatísticas Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Raiz de Roy foram altamente significativas. Nota-se que a interação Genótipo $\times$ Anos, que testa a hipótese de paralelismo, também foi rejeitada. Verifica-se também que os resultados foram os mesmos obtidos pelos testes univariados.

Tabela 13 - Resultado da Análise Multivariada Usando o proc GLM

H = Matriz da soma de quadrados e produtos cruzados para o fator Anos					
E = Matriz da soma de quadrados e produtos cruzados do resíduo					
S=1 $m_1 = 1$ $m_2 = 26$					
Estatística	Valor	F	Num. G.L. ¹	Den. G.L. ²	Pr > F
Lambda de Wilks	0,0194	680,39	4	54	< 0,0001
Traço de Pillai	0,9805	680,39	4	54	< 0,0001
Traço de Hotelling-Lawley	50,3994	680,39	4	54	< 0,0001
Raiz de Roy	50,3994	680,39	4	54	< 0,0001
H = Matriz da soma de quadrados e produtos cruzados para o fator Genótipo × Anos					
E = Matriz da soma de quadrados e produtos cruzados do resíduo					
S=4 $m_1 = 7$ $m_2 = 26$					
Lambda de Wilks	0,0066	7,30	76	215,1	< 0,0001
Traço de Pillai	2,7044	6,26	76	228	< 0,0001
Traço de Hotelling-Lawley	12,3923	8,59	54	159,88	< 0,0001
Raiz de Roy	7,0474	21,14	19	57	< 0,0001

1: Grau de liberdade do numerador; 2: Grau de liberdade do denominador.

5.8 Análise Via Modelo Misto

Para selecionar a estrutura da Matriz de Covariâncias foram testados oito modelos, descritos na Tabela 14.

Considerando o critério de AIC a estrutura selecionada seria a UN, por apresentar o menor valor de AIC, porém conforme ocorreu na análise do experimento anterior, a estrutura UN apresenta o maior número de parâmetro (15 parâmetros) a ser estimado, o que é inviável. Nesse caso, uma segunda opção seria a estrutura ARH (1), com seis parâmetros a serem estimados, e com o menor valor de BIC e o segundo menor valor de AIC.

Tabela 14 - Estrutura da Matriz de Covariâncias, Parâmetros Estimados e Critérios de Informação de Akaike (AIC) e de Schwarz (BIC)

Estruturas da Matriz de Covariâncias	Parâmetros	AIC (Akaike)	BIC (Schwarz)
Simétrica Composta : CS	2	2147,1	2151,9
Não Estruturada: UN	15	2113,0	2148,8
Auto-Regressiva 1ª Ordem: AR (1)	2	2141,6	2146,4
Componentes de Variância: VC	1	2145,8	2148,2
Toeplitz: TOEP	5	2144,2	2156,1
Huynh-Feldt: HF	6	2143,8	2158,1
Simétrica Composta Heterogênea: CSH	6	2132,0	2146,3
Auto-Regressiva Heterogênea: ARH (1)	6	2124,1	2138,4

A Tabela 15 apresenta os testes para os efeitos fixos do modelo quando a estrutura da matriz de covariâncias é do tipo ARH (1).

Tabela 15 - Teste para os Efeitos Fixos do Modelo com a Estrutura da Matriz de Covariâncias do Tipo Auto-Regressiva Heterogênea

Causas de Variação	Num. G.L.	Den. G.L.	F	Pr > F
Bloco	3	57	2,90	0,0429
Genótipo	19	57	21,95	< 0,0001
Anos	4	240	529,81	< 0,0001
Genótipo × Anos	76	240	68,27	< 0,0001

Nota-se que os efeitos Genótipos, Anos e interação Genótipos × Anos foram significativos, portanto, devem permanecer no modelo.

Contudo, o perfil médio de resposta, em função da produtividade dos 20 genótipos no período de cinco anos, pode ser representado graficamente conforme a Figura 2.

Figura 2 - Perfis médios de TCH dos 20 genótipos durante cinco anos

A representação gráfica (Figura 2), mostra que os perfis médios de produtividade dos genótipos em função dos anos, mais uma vez, fortalece os resultados, constatando que as hipóteses dos perfis horizontais, paralelos e coincidentes, foram rejeitadas.

6 CONCLUSÕES

De acordo com os objetivos deste estudo, pode-se concluir:

- Não houve diferença entre os resultados dos testes para as diferentes metodologias;
- Em relação a análise univariada, nos dois experimentos avaliados, o teste de esfericidade foi rejeitado, porém a significância dos testes para os fatores intra-indivíduos não foi alterada com as correções dos números de graus de liberdade;
- Em relação a análise multivariada, os quatros testes empregados foram altamente significativos para os fatores Anos e a interação Genótipo $\times$ Anos, portanto, os resultados foram os mesmos obtidos pelos testes univariados.
- Utilizando-se a metodologia de modelo misto, observou-se que, segundo o critério de AIC, nos dois experimentos, o modelo da matriz de covariâncias apropriado é o Não Estruturado (UN), porém esse modelo apresentou o maior número de parâmetros a ser estimado. Como segunda opção, pelo critério de BIC, têm-se os modelos Simétrica Composta Heterogênea (CSH), para o Experimento 1 e o modelo Auto-Regressiva Heterogênea (ARH (1)), para o Experimento 2.

Para continuidade deste estudo, o modelo misto é uma alternativa bem interessante para melhoramento genético da cana-de-açúcar, especialmente durante a seleção dos melhores genótipos, pois dada sua flexibilidade e precisão, será possível obter estimativas mais seguras dos componentes de variância e prever os valores genotípicos, que por fim poderão proporcionar a previsão de produção de uma futura colheita para um determinado genótipo.

REFERÊNCIAS

- AUBIN, E.C.Q. **Análise de experimentos com medidas repetidas**. 1984. 164p. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1984.
- BARBOSA, G.V.S. Interpretação de resultados experimentais na agroindústria da cana-de-açúcar. In: SIMPÓSIO DA AGROINDÚSTRIA DA CANA-DE-AÇÚCAR DE ALAGOAS, 22., 2005, Maceió. **Anais...** Maceió: STAB, 2005. 1CD-ROM.
- BOX, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems I. Effects of inequality of variance in the one-way classification. **Annals of the Mathematical Statistics**, Ann Arbor, v.25, p.290-302, 1954.
- 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.
- BÍBLIA sagrada. 70.ed. São Paulo: Ave Maria, 1989. p. 997-1043.
- COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento da safra brasileira cana-de-açúcar safra 2007/2008**. Brasília, 2007.
- COSTA, S.C. **Modelos lineares generalizados para dados longitudinais**. 2003. 110 p. Tese (Doutorado em Estatística)- Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.
- CROWDER, M.J.; HAND, D.J. **Analysis of repeated measures**. London: Chapman e Hall, 1990, 256p.
- DIGGLE, P.J. An approach to the analysis of repeated measurements. **Biometrics**, Washington, v.44, p.959-971, 1988.
- DIGGLE, P.J.; LIANG, K-Y.; ZEGER, S.L. **Analysis of longitudinal data**. 4th.ed New York : Oxford University Press, 1998, 253p. (Oxford Statistical Science Series, 13).
- FERNANDEZ, G.C.J. Repeated measure analysis of line-source sprinkler experiments. **HortScience**, St. Joseph, v.26, n.4, p.339-342, 1991.
- FERREIRA, D.F. **Análise multivariada**. Lavras: Universidade Federal de Lavras; Departamento de Ciências Exatas, 1996. p.400. 1996.
- GILL, J.L. **Design and analysis of experiments in the animal and medical sciences**. Ames: The Iowa State University Press, 1986. 301p.
- GUIMARÃES, P.R.B. **Modelo linear misto de Laird-Ware**: predição de efeitos aleatórios e estimação de parâmetros via filtro de Kalman. Campinas, 1994. 135p. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, 1994.
- HENDERSON, C.R. Estimation of changes in herd environment. **Journal of Dairy Science**, Champaign, v. 32, p.706-711, 1949.

HENDERSON, C.R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, Washington, v.31, n.2. p.423-447. Jun. 1975.

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

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

HUYNH, H.; FELDT, L.S. Some approximate tests for repeated measurement designs. **Psychometrika**, Baltimore, v.43, n.2, p. 161-175, Jun. 1978.

KSHIRSAGAR, A.M.; SMITH, W.B. **Growth curves**. New York: Marcel Dekker, 1995, 359p.

KUEHL, R.D. **Statistical principles of research design and analysis**. Belmont: Duxbury Press, 1994. 686p.

KIRK, R.E. Randomized block designs. In:_____. **Experimental design: procedures for the behavioral**. Belmont: Wadsworth. 1995, chap.6, p.243-267.

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

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

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

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

MALHEIROS, E.B. 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.

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, p.957-965, 1991.

MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of messy data**. New York: Chapman & Hall, 1992. 473p.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3th ed. New York: McGraw-Hill, 1974. 564p.

- MORRISON, D.F. **Multivariate statistical methods**. 3th ed. New York: McGraw- Hill, 1990. 415p.
- MULLER, K. E. e BARTON, C. N. Approximate power for repeated-measures ANOVA lacking sphericity. **Journal of the American Statistical Association**, Boston, v.84, n.406, p.549-555, Jun 1989.
- REZENDE, D.M.L.C.; FERREIRA, D.F.; MINIZ, J.A. Comparações de técnicas de análises de experimentos utilizando medidas repetidas no tempo. **Ciência e Agrotecnologia**, Lavras, v.23, n.4, p.928-938, 1999.
- RIBOLDI, J.; FERNANDEZ, D.W.X.; CASTRO, S.M.J. Análise de observações simultâneas e medidas repetidas. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 41., 1996, São José do Rio Preto. **Resumos...** São José do Rio Preto: UNESP, 1996.
- RIBOLDI, J. Modelos mistos aplicados. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 52.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12., 2007. Santa Maria. **Anais...**, Santa Maria: SEAGRO, 2007.
- ROUANET, H.; LEPINE, D. Comparison Between Treatments in a Repeated-Measurement Design: ANOVA and Multivariate Methods. **British Journal Mathematical and Statistical Psychology**. London, v.38, n.2, p.148-150, 1970.
- ROSÁRIO, M.F. **Emprego do conceito de medidas repetidas na avaliação do desempenho de genótipos de frangos de corte**. 2003. 66p. Dissertação (Mestrado em Genética e Melhoramento de Plantas)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2003.
- ROSÁRIO, M.F.; SILVA, M.A.N.; SAVINO, V.J.M.; COELHO, A.A.D.; MORAES, M.C. Avaliação do desempenho zootécnico de genótipos de frangos de corte utilizando-se a análise de medidas repetidas. **Revista Brasileira de Zootecnia**, Viçosa, v.34, n.6, p.2253-2261, 2005.
- SAS INSTITUTE. SAS/STAT software: changes and enhancements through release 9.1. Cary, 2003. 5136p.
- SINGER, J.M.; ANDRADE, D.F. Análise de dados longitudinais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7., 1986. Campinas. **Anais...** São Paulo: UNESP, 1986.
- TIMM, N. H. Multivariate analysis with applications in education and psychology. **Journal of the American Statistical Association**, Boston, v.81, n.395. p.863-864, Sept. 1986.
- VON ENDE, C.N. 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. chap.6, p. 113-137.

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

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, abr./jun. 2001.

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

ANEXOS

ANEXO A - Conjunto de dados do Experimento 1

Tabela 16 - Produtividade de Cana-de-açúcar (TCH) dos Genótipos com três Cortes

Parcela	Genótipo	Bloco	Cortes em anos		
			ano 1	ano 2	ano 3
1	RB835054	1	139,670	137,200	127,590
2	RB835054	2	127,200	130,330	123,650
3	RB835054	3	133,000	128,200	117,440
4	RB835486	1	104,530	114,670	107,490
5	RB835486	2	94,070	110,870	93,950
6	RB835486	3	113,930	119,800	107,310
7	RB855453	1	148,070	131,000	117,240
8	RB855453	2	141,000	128,070	100,690
9	RB855453	3	151,200	123,530	114,000
10	RB855465	1	117,800	115,470	107,590
11	RB855465	2	114,870	121,730	110,550
12	RB855465	3	121,330	116,800	102,400
13	RB925211	1	133,870	124,930	106,800
14	RB925211	2	127,470	120,600	109,910
15	RB925211	3	140,270	121,930	113,840
16	RB925345	1	147,800	121,330	121,630
17	RB925345	2	154,670	124,870	104,910
18	RB925345	3	134,000	131,130	126,190
19	RB945961	1	134,330	118,930	88,310
20	RB945961	2	124,130	116,730	99,120
21	RB945961	3	129,870	116,130	95,400
22	RB946903	1	101,600	129,400	113,350
23	RB946903	2	114,600	120,470	119,550
⋮	⋮	⋮	⋮	⋮	⋮
84	SP91-1049	3	141,470	130,670	118,670

ANEXO B - Conjunto de dados do Experimento 2

Tabela 17 - Produtividade de Cana-de-açúcar (TCH) dos Genótipos com cinco Cortes

Parcela	Genótipo	Bloco	Cortes em anos				
			ano 1	ano 2	ano 3	ano 4	ano 5
1	SP791011	1	107,143	99,206	119,048	103,175	71,429
2	SP791011	2	97,222	85,317	111,111	101,190	85,317
3	SP791011	3	107,143	89,286	113,095	105,159	79,365
4	SP791011	4	105,159	75,397	111,111	87,302	83,333
5	RB83594	1	105,159	47,619	105,159	53,571	47,619
6	RB83594	2	109,127	51,587	115,079	67,460	65,476
7	RB83594	3	103,175	51,587	105,159	51,587	55,556
8	RB83594	4	97,222	49,603	103,175	71,429	59,524
9	RB92579	1	134,921	83,333	125,000	85,317	77,381
10	RB92579	2	134,921	99,206	119,048	119,048	101,190
11	RB92579	3	138,889	79,365	115,079	130,952	115,079
12	RB92579	4	126,984	87,302	136,905	101,190	93,254
13	RB9364	1	111,111	79,365	101,190	65,476	59,524
14	RB9364	2	109,127	77,381	93,254	75,397	69,444
15	RB9364	3	103,175	83,333	99,206	71,429	73,413
16	RB9364	4	111,111	75,397	89,286	65,476	75,397
17	RB93509	1	99,206	93,254	119,048	89,286	81,349
18	RB93509	2	97,222	91,270	113,095	105,159	91,270
19	RB93509	3	103,175	89,286	138,889	81,349	85,317
20	RB93509	4	99,206	85,317	128,968	73,413	77,381
21	RB931003	1	113,095	103,175	138,889	113,095	73,413
22	RB931003	2	113,095	107,143	126,984	119,048	67,460
23	RB931003	3	107,143	93,254	128,968	125,000	85,317
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	RB941570	4	95,238	83,333	109,127	65,476	43,651

ANEXO C - Programa utilizado nas análises dos modelos univariado, multivariado e misto

```

data multi (keep = parcela genotipo bloco ano1 ano2 ano3 ano4 ano5)
    uni (keep = parcela genotipo bloco ano TCH);
    input parcela genotipo$ bloco ano1 ano2 ano3 ano4 ano5@@;
output multi;
TCH = ano1; ano = 1; output uni;
TCH = ano2; ano = 2; output uni;
TCH = ano3; ano = 3; output uni;
TCH = ano4; ano = 4; output uni;
TCH = ano5; ano = 5; output uni;
cards;
1   SP791011    1   107.143 99.206  119.048 103.175 71.429
2   SP791011    2   97.222  85.317  111.111 101.190 85.317
3   SP791011    3   107.143 89.286  113.095 105.159 79.365
...   ...      ...   ...      ...      ...      ...
78  RB941570    2   105.159 71.429  109.127 75.397  57.540
79  RB941570    3   103.175 83.333  126.984 71.429  53.571
80  RB941570    4   95.238  83.333  109.127 65.476  43.651
;
proc print;
run;
proc glm data=uni;
class genotipo bloco ano parcela;
model TCH = genotipo bloco bloco*genotipo ano ano*genotipo;
test h=genotipo e=bloco*genotipo;
run;
proc glm data=multi;
class bloco genotipo;
model ano1 ano2 ano3 ano4 ano5 = bloco genotipo/nouni;

```

```

manova h=genotipo/printe;
repeated ano 5 polynomial/printe summary;
lsmeans genotipo;
run;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=hf subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=un(1) subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=cs subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=vc subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=ar(1) subject=parcela(genotipo) r;
run ;

proc mixed data=uni;

```

```
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=toep subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=toeph subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH =bloco genotipo|ano;
repeated ano / type=csh subject=parcela(genotipo) r;
run ;

proc mixed data=uni;
class bloco genotipo ano parcela;
model TCH = bloco genotipo|ano;
repeated ano / type=arh(1) subject=parcela(genotipo) r;
run ;
```