MODELO UNIVARIADO APLICADO A DADOS LONGITUDINAIS DE CANA-DE-AÇÚCAR

Edjane Gonçalves de FREITAS[1]
Decio BARBIN[1]
Geraldo Veríssimo de Souza BARBOSA[2]
Monalisa Sampaio CARNEIRO[3]
Antônio Ismael BASSINELLO[3]

▪ RESUMO: Na fase experimental de um programa de melhoramento genético de cana-de-açúcar, são implantados vários experimentos. O interesse do pesquisador é avaliar o perfil produtivo dos genótipos (clones e variedades) em diferentes ocasiões (anos). Entre os pesquisadores de ciências agrárias é freqüente o uso do esquema de parcela subdividida para análise de dados longitudinais, devido a facilidade de análise e interpretação dos resultados. É ignorada a existência de alguma forma de relação entre as observações tomadas nas diferentes ocasiões, e a análise é realizada considerando o modelo univariado no esquema de parcela subdividida, que impõe forte restrição quanto à matriz de variância-covariância dos dados. O objetivo desse estudo é verificar se existem diferenças entre os resultados do teste F em relação a subparcela (Anos e interação Genótipos × Anos) quando a análise é realizada com e sem a correção dos números de graus de liberdade. Ressaltando uma forma mais segura e correta de aplicação do esquema de parcela subdividida, destacando suas vantagens e contribuições. Verificou-se que a significância dos testes para os fatores intra-indivíduos não foi alterada após as correções dos números de graus de liberdade. Entretanto, a aplicação dessa metodologia está sujeita a verificação da condição de esfericidade. Esse método mostrou-se eficiente, sendo uma alternativa ao uso de metodologias mais complexas, como as técnicas multivariadas e as análises via modelos mistos, que exigem um maior embasamento estatístico do pesquisador.

[1]Departamento de Ciências Exatas, Escola Superior de Agricultura "Luiz de Queiroz" – ESALQ, Universidade de São Paulo – USP, CEP: 13418-900, Piracicaba, São Paulo, Brasil. E-mail: *efreitas@esalq.usp.br* / *debarbin@esalq.usp.br*

[2]Departamento de Fitotecnia, Universidade Federal de Alagoas – UFAL, Campus Delza Gitaí, BR 104 Norte, Km 85, CEP: 57100-000, Rio Largo, Al, Brasil. E-mail: *gvsb@fapeal.br*

[3]Departamento de Biotecnologia Vegetal, Universidade Federal de São Carlos – UFSCar, Caixa Postal 173, CEP: 13600-970, Araras, SP, Brasil. E-mail: *monalisa@cca.ufscar.br* / *bassinello@cca.ufscar.br*

▪ PALAVRAS-CHAVE: Medidas repetidas no tempo; melhorameto genético; cana-de-açúcar; modelo univariado; parcela subdividida.

1 Introdução

Na fase experimental de um programa de melhoramento genético de cana-de-açúcar, são implantados vários experimentos. Nessa etapa, o interesse do pesquisador é avaliar o perfil produtivo dos genótipos (clones e variedades) em diferentes ocasiões (anos), com o objetivo de assegurar a seleção dos melhores genótipos. Estudo dessa natureza é bastante comum em várias áreas da pesquisa como: agricultura, sivicultura, medicina, economia e outras. Os planejamentos logitudinais são também chamados de análise de medidas repetidas no tempo, e ocorrem nos casos em que as unidades experimentais ou unidade de observação são sujeitas às medidas repetidas não aleatorizadas no tempo. Por exemplo, considerando um experimento para avaliação da produtividade agrícola de genótipos de cana-de-açúcar, medidas da produção são tomadas nas mesmas unidades experimentais durante três ou mais anos sucessivos, então, devido a forma sistemática em que as observações são tomadas, acredita-se que possa existir uma forma de relação entre as observações tomadas ao longo do tempo, e para que a análise dos dados seja correta e produza resultados confiáveis deve-se admitir outras suposições para modelo, além das usuais.

Do ponto de vista prático, a colheita da cana-de-açúcar se dá em anos sucessivos, na mesma parcela, assim como da grande maioria das culturas semi-perenes, mas não há muito interesse em analisar isso como medidas repetidas e sim cada colheita e a soma (ou média) de todas as colheitas. Entretanto, estatisticamente, a pressuposição de independência dos erros a nível de subparcela não é atendida, e a aplicação correta da metodologia proposta é fundamental, pois o seu resultado dará suporte a realização de outras análises, tal como, a verificação de correlação genética do caráter de um corte para outro, no conjunto dos clones. Vencovsky e Barriga (1992) avaliaram o redimento de açúcar de 33 clones de cana-de-açúcar durante três anos, mostrando a correlação entre os cortes. Segundo os autores, é possível avaliar o redimento esperado de um genótipo num dado corte por meio da seleção em outros cortes, ou seja, trata-se de uma seleção indireta. Assim, é possível verificar a contribuição da seleção em diferentes cortes.

Entre os pesquisadores de ciências agrárias é frequente o uso do esquema de parcela subdividida para análise de dados longitudinais, devido a fatores como a facilidade de análise e de interpretação dos resultados. Neste caso, é ignorada a existência de alguma forma de relação entre as observações tomadas nas diferentes ocasiões (tempo), e a análise é realizada considerando o modelo univariado com o esquema de parcela subdividida, que impõe uma forte restrição quanto à matriz de variância-covariância dos dados nos diferentes tempos.

Malheiros (2004) também afirma que a análise de experimentos com medidas repetidas no tempo usando o esquema em parcelas subdivididas, tendo o tempo como subparcelas, é prática comum, mas nem sempre correta, pois, como se sabe, esse esquema pressupõe que a matriz de covariâncias Σ tenha uma estrutura homogênea, o que nem sempre é verificado.

Em estudo de medidas repetidas, a estrutura de parcelas subdivididas, segundo Xavier e Dias (2001) é caracterizada quando se aplicam às parcelas os níveis de um 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. Enquanto que, em estudos de medidas repetidas no tempo, com o 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.

Há semelhanças entre os experimentos com medidas repetidas no tempo e aqueles em parcelas subdivididas, onde 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. Assim, as respostas de diferentes subparcelas, na mesma parcela, são igualmente correlacionadas umas com as outras. Já em experimentos com medidas repetidas no tempo, as respostas de tempos mais próximos são, em geral, mais fortemente correlacionadas do que as de tempos mais distantes (Littell et al. 1998; Costa, 2003; Rosário, 2003).

Logo, quando o pesquisador analisa um experimento com medidas repetidas no tempo utilizando o esquema de parcela subdividida, estará violando duas pressuposições básica requeridas pela análise de variância, a falta de casualização dos efeitos de tempos (anos) e da interação genótipos (tratamentos) $\times$ anos e a dependência dos erros, já que as observações foram tomadas nas mesmas unidades experimentais ao longo do tempo. Além disso, foi demonstrado por Huynh e Feldt (1970), que 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.

Alguns autores como 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.

Uma condição suficiente para que o teste F da análise de variância usual, em nível de subparcela, para o fator anos e interação tratamentos $\times$ anos, seja válido, é que a matriz de covariâncias tenha uma forma chamada de simétrica composta (Xavier, 2000). Isso ocorrerá quando a matriz de covariâncias Σ , puder ser expressa, da seguinte forma:

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

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 simetrica 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 tem a forma de simétrica 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 (anos e interação tratamento $\times$ anos).

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.

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 (Huynh e Feldt, 1970; Rouanet e Lépine, 1970). Uma condição mais geral da forma de Σ, foi denominada de H-F, e é equivalente à especificar que as variâncias da diferença entre pares de erros sejam todas iguais, e portanto, essa condição corresponde a matriz simétrica composta.

Calculando-se todas as variâncias das diferenças dos possíveis pares de erros pode-se comprovar por:

$$\sigma^2_{Y_j - Y_{j'}} {=} \sigma^2_{Y_j} + \sigma^2_{j'} - 2\sigma^2_{jj'} = 2\lambda,$$

é constante para todo j e j' ($j \neq j'$), isto é, desde que $\sigma^2_{Y_j - Y_{j'}}$ 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.

Um problema que surge é quando a matriz de covariância não é do tipo simétrica composta, erros independentes e a condição de esfericidade não é satisfeita, o que leva a teste F não exato. Esse problema pode ser contornado utilizando-se o teste de esfericidade proposto por Mauchly (1940), e caso a condição de esfericidade seja rejeitada, para validar a análise dos dados de medidas repetidas com a aplicação de análise univariada no esquema de parcelas subdivididas é necessário realizar o ajuste do número de graus de liberdade do teste F para os efeitos intra-indivíduos (subparcela).

1.1 Teste de Mauchly

Para verificar a validade da condição de esfericidade da matriz de covariâncias, Mauchly (1940), desenvolveu um critério que testa se uma população normal multivariada tem variâncias iguais e correlações nulas, sendo chamada de "esférica", quando tiver essa simetria.

A estatística de teste formulada por Mauchly (1940) para verificação da condição de esfericidade 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 constrastes 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 pela 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 $-\gamma ln(\mathbf{W}) > \chi^2_{a,f}$, rejeita-se a hipótese de nulidade ao nível a 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.

1.2 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.

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. Geisser e Greenhouse (1958); Huynh e Feldt (1976) também sugeriram ajustes para o número de graus de liberdade do teste F para o fator erro intra-indivíduos. 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: $A_{q\times q} = \mathbf{CSC}'$, 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, os testes F's para os fatores anos e interação genótipo $\times$ anos 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 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.

Muller e Barton (1989), 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 também em situações onde a matriz de covariâncias é diferente de tratamento para tratamento. 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 univarida, 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.

Trabalhando com dados simulados, Malheiros (2004) concluiu que tanto para efeito de tempo como para a interação tratamento $\times$ tempo, com a correção de Huynh e Feldt (1976) os testes F da análise da variância foram imprecisos e com a correção de Greenhouse e Geisser (1959) foram precisos, independente dos dados serem balanceados ou não e da estrutura da matriz Σ. Esse fato foi evidenciado anteriormente por Littell et al.(1998), que sugeriram a correção de Greenhouse e Geisser (1959).

Utilizando-se o modelo univariado no esquema de parcelas subdivididas proposto por Vonesh e Chinchilli (1997), com o delineamento de blocos ao acaso, têm-se:

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

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, que, nesse estudo, refere-se a anos,$(\beta\tau)_{ij}$ 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 $i = 1,...,$ b, $j = 1,..., g$ e $k = 1,..., t$.

Nesse modelo, 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.

1.2.1 Teste de hipótese sobre os parâmetros do modelo

Através do método de análise univariada "*split-plot*", é testado se existe diferença entre genótipos, entre as condições de avaliação (anos) e entre as interações (genótipos $\times$ anos). Para isso, serão testadas as seguintes hipóteses de interesse:

1) Para efeito de genótipos

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

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

2) Para efeito de anos

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

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

3) Para efeito da interação (genótipos × anos)

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

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

Aplicando-se o modelo univariado com esquema de parcela subdivida, buscou-se nesse trabalho verificar se existem diferenças entre os resultados do teste F em relação a subparcela (Anos e interação Genótipos × Anos) quando a análise é realizada com e sem a correção dos números de graus de liberdade, caso a condição de esfericidade não seja satisfeita. E também ressaltar uma forma mais segura e correta de aplicação do esquema de parcela subdividida para a análise de experimentos de cana-de-açúcar com medidas repetidas ao longo do tempo. Destacando, as vatagens e contribuições da metodologia, auxiliando o pesquisador a obter resultados seguros por meio de técnicas mais simples.

2 Material e métodos

Foram considerados dados de um experimento de cana-de-açúcar da Rede Interuniversitária de Desenvolvimento do Setor Sucroalcooleiro-RIDESA, conduzido pelo Programa de Melhoramento de Genético da Cana-de-açúcar, PMGCA-UFSCar da Universidade Federal de São Carlos. Instalado na Usina Itamarati, em São Paulo, em 05/05/2003, com 28 genótipos, no delineameto 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, com dados de cana-planta, soca e ressoca. A variável analisada foi TCH - Tonelada de Cana por Hectare, medida sempre na mesma unidade experimental durante três anos sucessivos. Considerando-se a metodologia de dados longitudinais (análise de medidas repetidas), para analises dos dados, adotou-se o modelo univariado com esquema de parcela subdividida, sugerido por Vonesh e Chinchilli (1997), conforme já comentado. Utilizando-se um procedimento do SAS, o proc GLM, realizou-se a análise univariada, e após, verificou-se através do teste de Mauchly (1940), se a condição de esfericidade foi atendida, caso seja rejeitada, realizar-se-á nova análise com correção dos números de graus de liberdade. A representação do esquema de análise pode ser entendida conforme a Figura 1, que segue:

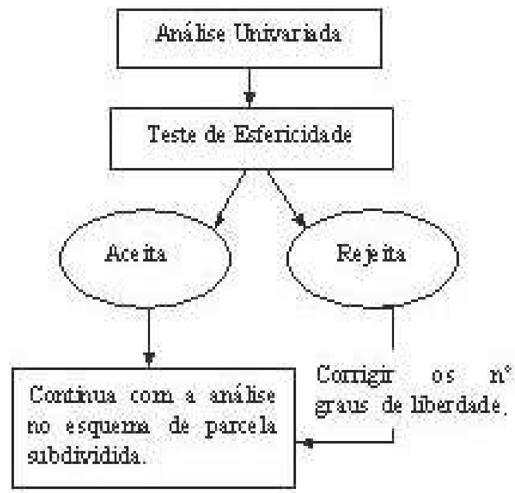

Figura 1 - Diagrama para análise univarida com teste de esferidade e correção dos números de graus de liberdade.

3 Resultados e discussão

Conforme a metodologia proposta, foram realizadas as análises do experimento, obtendo-se os resultados que seguem:

3.1 Análise Univariada

As Tabelas 1 e 2 apresentam os resultados da análise univariada e teste de esfericidade, via proc GLM, no esquema de parcelas subdivididas.

Observa-se que existem diferenças siginficativas dos fatores Genótipo, Anos e da interação Genótipo × 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. Verifica-se ainda que, as estatísticas $R^2 = 0,9491$ e C.V. = 4,58%, indicam um bom ajuste do modelo e alta precisão experimental, 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 1.1.

Pode-se observar na Tabela 2, 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 Box (1954); Geisser e Greenhouse (1958); Huynh e Feldt (1976), é necessário realizar a correção para os números de graus de liberdade dos fatores intra-indivíduos, nesse caso, para o fator Anos e a interação Genótipo x Anos.

Tabela 1 - Resultado da análise univariada usando o proc GLM

Variável Dependende: TCH				
Causas de Variação	**G.L.**	**Q.M.**	**F**	**Pr > F**
Modelo	139	463,31632	15,02	$< 0,0001$
Resíduo	112	30,85189		
Total Corrigido	251			
Bloco	2	77,37318	1,45	0,0860
Genótipo	27	1357,97647	25,37	$< 0,0001$
Genótipo × Bloco	54	53,53318	1,74	0,0074
Anos	2	6699,92601	217,16	$< 0,0001$
Genótipo × Anos	54	209,07804	6,78	$< 0,0001$
$R^2 = 0,9491$ C.V. = 4,58 % Média de TCH = 121,279				

Tabela 2 - Teste de Esfericidade

Hipótese	**Teste de Mauchly(W)**	**Aprox. Qui-Q.**	**Pr.**$> \chi^2_{(0,05;2)}$
H_0:CΣC´$= \lambda$I	0,7870	12,6927	0,0018

3.2 Análise univariada e correção dos números de graus de liberdade

Na Tabela 3 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 de produtividade dos genótipos são diferentes (p<0,0001). Já hipótese de perfis paralelos, pode ser verificada através do teste para a interação Genótipo x 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 fornecida 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 estatístiscas Greenhouse-Geisser ($\hat{\varepsilon} = 0,8244$) e Huynh-Feldt ($\tilde{\varepsilon} = 1,3037$) foram considerados. 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.

No presente trabalho, a utilização da metodologia univariada para análise de dados logitudinais de cana-de-açúcar é recomendada, pois aliada as vantagens de facilidade de aplicação e interpretação dos resultados, mostrou-se ser uma ferramenta eficiente para análise de dados dessa natureza. Neste caso, é possível dispensar o uso de metodologias mais complexas, como as técnicas multivariadas e as análises via modelos mistos, que exigem do pesquisador um maior colhecimento estatístico. Além dessas vantagens, contribui para obtenção de estimativas seguras

Tabela 3 - 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.**	**Q.M.**	**F**	**Pr > F**		
Bloco	2	77,37318	1,45	0,2446		
Genótipo	27	1357,97647	25,37	<0,0001		
Resíduo	54	53,53318				
TESTE PARA OS FATORES INTRA INDIVÍDUOS						
C.V.	**G.L.**	**Q.M.**	**F**	**Pr > F**	**G-G**	**H-F**
Anos	2	6699,926	215,32	<0,0001	<**0,0001**	<**0,0001**
Anos × Bl.	4	23,723	0,76	0,5519	0,5294	0,5519
Gen. × Anos	54	209,078	6,72	<0,0001	<**0,0001**	<**0,0001**
Res. (Ano)	108	31,116				
Greenhouse-Geisser $\hat{\varepsilon} = 0,8244$						
Huynh-Feldt $\tilde{\varepsilon} = 1,3037$						

de componentes genéticos e fenotípicos que fornecem as estimativas do coeficiente de hedabilidade e do ganho de seleção (dados não demonstrados).

Embora nesse estudo, não tenha sido evidenciado diferenças dos resultados nas análises com e sem a correção do número de graus de liberdade para os fatores de subparcelas, é importante ressaltar que trata-se apenas de um caso específico, e esse ressultado não dispensa a verificação da condição de esfericidade da matriz de variância-covariância.

Conclusões

A análise estatistica dos dados do referido experimento permitiu as seguintes conclusões:

a) A significância dos testes para os fatores intra-indivíduos não foi alterada após as correções dos números de graus de liberdade. Porém, isso não descarta a importância de se realizar as correções dos números de graus de liberdade, uma vez que aplicação correta dessa metodologia, contribuirá para obtenção de estimativas mais seguras de componentes genéticos e fenotípicos que fornecem as estimativas do coeficiente de hedabilidade e do ganho de seleção.

b) O modelo univariado com esquema de parcela subdividida pode ser utilizado com vantagem para análise de dados longitudinais em experimentos com cana-de-açúcar, pois além da facilidade de aplicação e interpretação dos resultados, mostrou-se ser uma alternativa eficiente ao uso de metodologias mais complexas, como as técnicas multivariadas e as análises via modelos mistos, que exigem um maior embasamento estatístico do pesquisador. Porém, sua aplicação está sujeita a verificação da condição de esfericidade.

c) Para um programa de melhoramento de cana-de-açúcar que possui um histórico de dados provinientes de estudos ao longo do tempo, o método univariado consiste em uma ferramenta de análise simples, facilitadora e que pode ser utilizada de modo preciso e consistente por pesquisadores da área, na avaliação de experimentos de cana-de-açúcar com diferentes colheitas, e assim, de forma segura é possível formar o perfil produtivo dos genótipos que auxiliará na seleção dos melhores clones.

Agradecimentos

Este trabalho é parte da dissertação de mestrado do primeiro autor, desenvolvido no Departamento de Ciências Exatas da ESALQ/USP com o apoio do Programa de Melhoramento Genético da Cana-de-açúcar da Universidade Federal de São Carlos – PMGC/UFSCar.

FREITAS, E. G; BARBIN, D.; BARBOSA, G. V. S.; CARNEIRO, M. S.; BASSINELLO, A. I. Univariado model applied data longitudinal with sugarcane. *Rev. Bras. Biom.,* São Paulo, v.26, n.2, p.93-106, 2007.

▪ *ABSTRACT: In the test phase of a program for genetic improvement of sugarcane, are implanted several experiments. The interest of the search is assess the profile of the productive genotypes (clones and varieties) on different occasions (Years). Among researchers of agricultural sciences is frequent use of the split plot designs for analysis of data longitudinal, due to ease of analysis and interpretation of the results. It ignored the existence of some form the relationship between the observations taken at different times, and the analysis is performed considering the univariate model with the split plot designs, which requires a strong restriction on the variance - covariance matrix of the data . The objective of this study is to verify whether there are differences between the results the F test in relation to the subplot (years and interaction genotypes × years) when the analysis is performed with and without the correction of the numbers of degrees of freedom (if the condition of sphericity not satisfied), highlighting a more safe and correct application of the split plot designs for the analysis of experiments of sugarcane. The results of the analysis showed that the significance the tests for intra-subjects factors was unchanged after the correction of figures the degrees of freedom.*

▪ *KEYWORDS: Repeated measures in time; genetic improvement; sugarcane; univariate model; split plot.*

Referências

BOX, G. E. P. Some theorems on quadratic forms apllied to the study of analysis of variance problems I, II. *Ann. Math. Stat.*, Ann Arbor, v.25, p.290-302, 1954.

COSTA, S. C. *Modelos lineares generalizados para dados longitudinais*. 2003. 110f. Tese (Doutorado em Estatística)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2003.

GEISSER, S.; GREENHOUSE, S. W. *An extension of Box's results on the use of the F distribution in multivariate analysis. Ann. Mathem. Stat.*, Ann Arbor, v.29, p. 855-91, 1958.

GILL, J. L. *Design and analysis of experiments in the animal and medical sciences.* Ames: The Yowa State University Press, 1986. 301p.

HUYNH, H.; FELDT, L. S. Condition under which mean square rations in repeated measurements designs have exact F-distributions. *J. Am. Stat. Assoc.*, Alexandria, v.65, p.1582-1589, 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. *J. Ed. Stat.*, v.1, n.1, p.69-82, 1976.

KIRK, R. E. *Experimental design*: procedures for the behavioral. Pacific Grove: Books/Cole Publishing Company, 1995. p.62-65.

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

LITTELL, R. C.; HENRY, P. R.; AMMERMAN, C. B. Statistical analysis of repeated measures data using SAS procedures. *Am. Soc. Anim. Sci.*, Savoy, 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. *Rev. Mat. Estat.*, São Paulo, v.22, n.2, p.23-29, 2004.

MULLER, K. E.; BARTON, C. N. Approximate power for repeated-measures ANOVA lacking sphericity. *J. Am. Stat. Assoc.*, New York, v.84, n.406, p.549-555, 1989.

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.

ROSÁRIO, M. F. *Emprego do conceito de medidas repetidas na avaliação do desempenho de genótipos de frangos de corte*. 2003. 66f. Dissertação (Mestrado em Genética e Melhoramento de Plantas)- Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2003.

ROUANET, H.; LEPINE, D. Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. *Brit. J. Math. Stat. Psic.*, Wales, v.23, p.147-163, 1970.

SAS INSTITUTE *SAS OnlineDoc® 9.1.3*. Cary, 2004.

VENCOVSKY, R.; BARRIGA, P. *Genética biométrica no fitomelhoramento.* Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.

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

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

Recebido em 14.12.2007.

Aprovado após revisão em 26.06.2008.