

Modelos longitudinais aplicados a dados de experimentos com cupuaçuzeiro no Estado do Pará

Vinicius Silva dos Santos - FacEst, UFPA¹²

Maria Regina Madruga - FacEst, UFPA¹

Rafael Moysés Alves - Embrapa, Amazônia Oriental¹

Resumo: Neste estudo foram considerados os dados de um experimento de melhoramento genético do cupuaçuzeiro, realizado no Estado do Pará, envolvendo 17 cultivares acompanhados durante 4 anos. Por conta da natureza longitudinal das observações utilizou-se técnicas de análise univariada e multivariada de perfis em dados com medidas repetidas, utilizando como variável resposta o diâmetro médio da planta de cupuaçuzeiro, a fim de verificar se ocorrem diferenças significativas entre os resultados dos testes. Concluiu-se que os resultados para os testes de efeitos fixos nas análises uni e multivariadas de perfis foram os mesmos, levando a rejeição dos perfis paralelos (interação cultivares x tempos), horizontais (tempos) e a não rejeição dos perfis coincidentes (tratamentos).

Palavras-chave: análise de perfis, teste de esfericidade, medidas repetidas.

1 Introdução

O cupuaçuzeiro é uma das fruteiras mais conhecidas da região amazônica, sendo o Estado do Pará o maior produtor nacional, com mais de 14.000 ha plantados. No programa de melhoramento genético dessa espécie, assim como nas demais espécies perenes, o objetivo principal é a obtenção de genótipos de alta produção e ao mesmo tempo, resistentes a doenças. Na cultura do cupuaçuzeiro a doença conhecida como vassoura-de-bruxa é o principal entrave, ver [6]. Porém, por apresentar longo ciclo reprodutivo e ocupar grandes áreas experimentais, o custeio da pesquisa torna-se extremamente elevado. Portanto, o uso correto de técnicas estatísticas é imprescindível para a eficiência do melhoramento, reduzindo, assim, o tempo necessário para o lançamento de uma nova cultivar.

A análise de dados longitudinais ou medidas repetidas no tempo se refere a observações realizadas na mesma unidade experimental (indivíduo, planta, empresa, máquina, etc.) em mais de uma ocasião. No estudo de dados longitudinais, devem-se levar em consideração as estruturas das matrizes de correlação e de variâncias-covariâncias das observações.

Uma das técnicas mais utilizadas para avaliar dados com medidas repetidas é a análise uni/multivariada de perfis que tem como objetivo estudar os perfis médios de respostas dos tratamentos nos diferentes tempos. Na análise univariada, o modelo é considerado num esquema de parcelas subdivididas, onde o efeito de tratamento é alocado às parcelas, testando-se a hipótese de perfis coincidentes. O efeito de tempo e a interação entre tratamento e tempo são alocados às subparcelas, onde são testadas as hipóteses de perfis horizontais e paralelos, respectivamente. Avaliar o modelo desta forma, muitas vezes, acaba sendo incorreto, pois duas principais pressuposições da análise de variância usual são violadas: a não casualização das parcelas (tratamentos) e a dependência dos erros, devido às medidas repetidas no tempo, ocasionando a correlação entre os dados, sendo que, a correlação é maior entre medidas de tempos mais contíguos, além do que, erros correlacionados podem possuir uma matriz de covariâncias heterogêneas. Assim, o teste F para o efeito de tratamentos

¹ Agradecimento à EMBRAPA Amazônia Oriental

² Contato: Vinicius_est@yahoo.com.br

será válido. Entretanto, com relação à subparcela, só terá validade se satisfizer uma das pressuposições, a de esfericidade, que consiste em considerar variâncias iguais e covariâncias nulas.

Uma alternativa para análise de dados longitudinais no delineamento de parcelas subdivididas quando da hipótese de esfericidade ou circularidade ser rejeitada, é o uso das estatísticas de Greenhouse-Geisser (ver [1]) ou Huynh-Feldt (ver [2]), que ajustam os graus de liberdade do teste F para o fator das subparcelas. Outra forma é utilizar a análise multivariada de perfis, também conhecida como análise multivariada de variância, que assume uma matriz de covariâncias sem estrutura, considerando variâncias e covariâncias diferentes para qualquer par de observações (mais detalhes em [5]).

O objetivo deste trabalho foi aplicar as técnicas de análise univariada e multivariada de perfis em dados com medidas repetidas, verificando se ocorrem diferenças entre os resultados dos testes.

2 Metodologia

Os dados utilizados foram de um experimento instalado em fevereiro de 1998 na base física da Embrapa Amazônia Oriental, em Belém-PA. O experimento consistiu em avaliar 17 progênies de cupuaçuzeiro, dispostas em cinco blocos casualizados, com cinco plantas por parcela. A variável reposta utilizada foi o diâmetro médio da planta, avaliada nos meses de março/2000, março/2001, março/2002 e maio/2003. Como os intervalos de tempo não são regulares, foram adotados como instantes de avaliação os meses 25, 37, 49 e 63. As análises foram feitas em nível de médias de parcelas, utilizando o procedimento GLM do software SAS. Com o objetivo de estudar o perfil médio das cultivares, foram utilizados dois modelos para análise de medidas repetidas no tempo: o Univariado e o Multivariado.

a. Modelo Univariado

Na análise de variância univariada, considerando-se um experimento em blocos ao acaso, num esquema de parcelas subdivididas no tempo (análise de dados longitudinais), foi proposto o seguinte modelo matemático (ver [7]):

$$Y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \gamma_k + (\beta\gamma)_{ik} + (\tau\gamma)_{jk} + e_{ijk} \quad (1)$$

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, $(\beta\tau)_{ij}$ é o efeito aleatório da “interação” do i -ésimo bloco com o j -ésimo tratamento (erro associado às parcelas), γ_k é o efeito do k -ésimo tempo observado, $(\beta\gamma)_{ik}$ é o efeito aleatório da “interação” do i -ésimo bloco com o k -ésimo tempo, $(\tau\gamma)_{jk}$ é o efeito da interação entre o j -ésimo tratamento e o k -ésimo tempo e e_{ijk} é o erro aleatório associado à observação do k -ésimo tempo para o j -ésimo tratamento no i -ésimo bloco (variação do acaso sobre as observações), supostos homocedásticos, independentes e normalmente distribuídos, com $i = 1, \dots, b$, $j = 1, \dots, g$ e $k = 1, \dots, t$. Segundo [3], os blocos são construídos apenas para controlar a heterogeneidade entre as unidades experimentais, e assim não se deve esperar que o termo de interação $(\beta\gamma)_{ik}$ em (1) seja significativo.

Para que o teste F da análise univariada de variância, em nível de subparcela (fator tempo e a interação tempo x tratamento) seja válido é necessário que a matriz de variâncias-covariâncias Σ satisfaça a condição de esfericidade, ou seja,

$$\sigma_{ij} = \frac{\sigma_i^2 + \sigma_j^2}{2} - \lambda \quad (2)$$

para todo $i \neq j$ e λ uma constante positiva.

A condição H-F pode ser verificada através de um teste chamado teste de esfericidade (ver [4]), que consiste em avaliar se uma população multivariada apresenta variâncias iguais e ausência de correlação. Se o teste de esfericidade resultar significativo, ou seja, a condição de esfericidade não for satisfeita, realiza-se uma análise univariada aproximada corrigindo os graus de liberdade do numerador e denominador dos testes F para os efeitos intra-indivíduos. Para obtenção dessas correções, considera-se s_{ij} a i -ésima linha e a j -ésima coluna da matriz de covariâncias amostral $S_{(t \times t)}$, para o erro experimental intra-indivíduos. Escolhem-se $q = (t-1)$ contrastes ortogonais normalizados, sobre t medidas repetidas, tomando a matriz $C_{(q \times t)}$ em que as linhas são contrastes ortogonais normalizados nas t medidas repetidas.

b. Modelo Multivariado

Quando a condição de esfericidade não é satisfeita, uma alternativa para avaliar dados longitudinais é a análise multivariada, que considera a estrutura de covariâncias do tipo sem estrutura. O seguinte modelo é sugerido por [7] para um experimento com delineamento de parcelas subdivididas:

$$Y_{ijk} = \mu + \beta_i + \tau_j + \gamma_k + (\tau\gamma)_{jk} + e_{ijk} \quad (3)$$

com as mesmas notações do modelo (1), excluindo as interações $(\beta\gamma)_{ik}$ e $(\beta\tau)_{ij}$, referente ao erro associado às parcelas.

Os testes multivariados utilizados por ordem de preferência em termos do poder foram: Traço de Pillai, Lambda de Wilks, Traço de Lawley-Hotelling e Raiz Máxima de Roy (ver [8]).

3 Resultados

Para o teste de esfericidade de Mauchly, obtidos da análise multivariada, observou-se o valor da estatística qui-quadrado aproximada igual a 39,4159 (p-valor < 0,0001), significativa ao nível de 5%, violando a suposição de esfericidade, ou seja, a matriz de covariâncias não pode ser considerada do tipo Huynh-Feldt (HF). Por conta disto, fez-se o uso das estatísticas Greenhouse-Geisser e Huynh-Feldt, que ajustaram os graus de liberdade dos fatores intra-indivíduos, neste caso, *mês* e a interação *mês x progênie*, rejeitando as hipóteses de perfis horizontais e paralelos, respectivamente, ou seja, ocorre um crescimento das distâncias médias em proporções diferentes ao longo do tempo. Na Tabela 1 observa-se, também, a não rejeição da hipótese de perfis coincidentes ao nível de significância de 5%, para o fator *progênie*, demonstrando a igualdade das distâncias médias dos tratamentos.

Os resultados para a Análise Multivariada usando o proc GLM estão listados na Tabela 2, nota-se que os quatro testes apresentaram os mesmos resultados para ambos os fatores avaliados, rejeitando a hipótese de nulidade do efeito *mês* e da interação *mês x progênie*. Vale observar que esses resultados foram semelhantes aos obtidos pelo modelo univariado, apresentado na Tabela 1.

Tabela 1: Análise univariada e correções para o número de graus de liberdade para os efeitos.

TESTE PARA OS FATORES ENTRE INDÍVIDUOS							
Causas de Variação	G.L.	S.Q.	Q.M.	F	Pr > F		
Bloco	2	377,1421	188,5711	1,38	0,2671		
Progênie	16	2992,5496	187,0343	1,37	0,2205		
Resíduo	32	4384,4107	137,0128				
TESTE PARA OS FATORES INTRA INDÍVIDUOS							
Causas de Variação	G.L.	S.Q.	Q.M.	F	Pr > F	G - G	H - F
Mês	3	115763,7047	38587,9016	3763,63	<0,0001	<0,0001	<0,0001
Mês x Bloco	6	144,9293	24,1549	2,36	0,0363	0,067	0,0363
Mês x Progênie	48	1319,6558	27,4928	2,68	<0,0001	0,0006	<0,0001
Resíduo (Mês)	96	984,2739	10,2529				

Tabela 2: Resultado da Análise Multivariada para os efeitos intra-indivíduos.

Matriz da soma de quadrados e produtos cruzados para o fator <i>mês</i>					
Estatística	Valor	F	Gl-num ¹	Gl-den ²	Pr > F
Lambda de Wilks	0,0043	2337,67	3	30	<,0001
Traço de Pillai	0,9957	2337,67	3	30	<,0001
Traço de Hotelling-Lawley	233,7672	2337,67	3	30	<,0001
Raiz de Roy	233,7672	2337,67	3	30	<,0001
H = Matriz da soma de quadrados e produtos cruzados para o fator <i>mês x progênie</i>					
Lambda de Wilks	0,1586	1,61	48	90,021	0,0265
Traço de Pillai	1,2877	1,50	48	96	0,0458
Traço de Hotelling-Lawley	2,8961	1,74	48	65,803	0,0183
Raiz de Roy	1,9886	3,98	16	32	0,0004

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

4 Conclusões

As análises uni e multivariada de perfis apresentaram os mesmos resultados para os efeitos fixos, sendo que os testes para os fatores intra-indivíduos mostraram-se altamente significativos, ou seja, para avaliar o crescimento de plantas de cupuaçuzeiro deve-se levar em conta o efeito de tempo e sua interação com as progênies.

Referências

- [1] GEISSER, S.; GREENHOUSE, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals Mathematical Statistics, United States*, v.29, n.3, p.855-91, 1958.
- [2] HUYNH, H.; FELDT, S. L. Conditions under which mean square ratios in repeated Measurements designs have exact F-distributions. *Journal American Statistical Association, United States*, v.65, 1582-1570.

- [3] LIMA, C. G. *Análise de dados longitudinais provenientes de experimentos em blocos casualizados*; 126p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", USP, 1996.
- [4] MAUCHLY, J.W. Significance test for sphericity of a normal n-variate distribution. *Annals Mathematical Statistics*, United States, v.11, n.2, p.204-209, 1940.
- [5] 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. 106p.
- [6] SOUZA, A. G. C. de, ALVES, R. M., SOUSA, N. R., SOUZA, M. G. de Domesticação e melhoramento de cupuaçuzeiro In: *Domesticação e melhoramento de espécies amazônicas*. 01 ed. Viçosa : Suprema Editora Ltda., 2009, v.01, p. 319-331, 2009.
- [7] VONESH, F. E.; CHINCHILLI, V.M. *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker. 1997. 560p.
- [8] XAVIER, L. H. *Modelos univariado e multivariado para análise de medidas repetidas e verificação da acurácia do modelo univariado por meio de simulação*; 91 p. Dissertação (Mestrado em Estatística) – Escola Superior de Agricultura "Luiz de Queiroz", USP, 2000.