

Análise de variância corrigida para a produção de matéria seca ao longo do tempo

Juracy Mendes Moreira^{1*}, Aurélio Ferreira Melo², Jose Marcelo de Oliveira³, Daniela Silva Ataides⁴, Marcelo Carlos Ribeiro⁵, Juliano Bortolini⁶

¹Prof. Ms. Faculdade Almeida Rodrigues (FAR), Rio Verde – Goiás, Brasil. E-mail: juracimendesmoreira@yahoo.com.br

²Prof. Faculdade Almeida Rodrigues (FAR), Rio Verde – Goiás, Brasil. E-mail: aurelioferreirameloi@hotmail.com

³Prof. Ms. Faculdade Almeida Rodrigues (FAR), Rio Verde – Goiás, Brasil. E-mail: oliveirademarcelo@hotmail.com

⁴Profa. Esp. Faculdade Almeida Rodrigues (FAR), Rio Verde – Goiás, Brasil. E-mail: danisila@uol.com.br

⁵Doutorando em Estatística Aplicada e Biometria (UFV), Viçosa – Minas Gerais, Brasil. E-mail: marcelocarlosribeiro@hotmail.com

⁶Prof. Dr. Departamento de Estatística, Universidade Federal de Mato Grosso, Boa Esperança – Cuiabá, Brasil. E-mail: julianobortolini@ufmt.br

*Autor para correspondência.

RESUMO. Técnicas de análises de experimentos que utilizam medidas repetidas ao longo do tempo devem considerar a estrutura de correlação entre tempos e dentro de tempos. Neste trabalho foi analisado a produção de matéria seca com o uso de nutrientes de alta e baixa concentração, no qual foi constatado por meio do teste de Mauchly que a matriz de covariância do modelo proposto, não satisfaz ao critério de esfericidade, não possuindo variâncias iguais e correlações nulas. Esse fato pode tornar inválidos os testes F da análise de variância aplicados as fontes de variações presentes na subparcela, portanto foi utilizado duas formas de corrigir os respectivos graus de liberdade, e com isso garantir que a distribuição F seja exata. No estudo da produção matéria seca, ficou constatado que indefere o uso dos nutrientes na produção da matéria seca e este fato foi confirmado pela análise clássica de variância, bem como na análise de variância corrigida.

Palavras chave: Matéria seca, parcela subdividida no tempo, teste de esfericidade de Mauchly

Corrected variance analysis for the production of dry matter generated over time

ABSTRACT. Technical analysis experiments using repeated measures over time should consider the correlation structure between time and within time. Dry matter production with the usage of high and low nutrient concentration was analyzed in this work. (It) was found through Mauchly test the covariance matrix of the model (which) does not satisfy the criterion sphericity and do not have equal variances and zero correlations. This fact may turn the F test invalid for the analysis of variance applied to the sources of variation present in the subplot, so it used two paths to fix the respective degrees of freedom, and thus ensure that the F distribution is accurate. In the study of dry matter production, was verified that the use of nutrients in the production of dry matter is not relevant and this fact was confirmed by classical analysis of variance and the corrected variance analysis.

Keywords: Dry matter, subdivided portion in time, sphericity test mauchly

Introdução

No Brasil muitos experimentos são conduzidos com o objetivo de estudar e avaliar distintas culturas agrícolas, pois é um país que possui extrema importância na produção de grãos. Sendo assim, é importante que distintos experimentos sejam conduzidos no intuito de alcançar êxito na

produção agrícola e, com isso, várias instituições de ensino possuem papel preponderante para atingir este objetivo. Para que esses resultados experimentais reflitam informações coerentes e fidedignas é necessário que se proceda a adequada modelagem estatística. É comum que muitos

experimentos sejam conduzidos ao longo do tempo e, assim, parcelas experimentais reproduzem mais de uma medida ao longo do tempo, sendo essa característica denominada na literatura estatística de medidas repetidas. A expressão “medidas repetidas” é utilizada para especificar múltiplas observações da mesma característica em uma ou mais variáveis respostas na mesma unidade experimental (Nobre & Singer, 2007). Os experimentos com medidas repetidas no tempo envolvem dois fatores: tratamentos e tempos, e são frequentes em diversas áreas do conhecimento. O objetivo principal desse tipo de experimento é examinar e comparar as tendências dos tratamentos ao longo do tempo. Isso pode envolver comparações entre tratamentos dentro de cada tempo, ou comparações de tempos dentro de cada tratamento. Dessa forma, o tratamento é o fator entre indivíduos (unidades experimentais) e o tempo o fator intra indivíduos.

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

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

Em planejamentos longitudinais, a variável resposta pode ser contínua ou discreta, como, por

exemplo, ganho de peso e número de folhas. As unidades experimentais podem ser constituídas de um ou mais fatores ou tratamentos e cada uma dessas unidades pode gerar unidades observacionais, em que cada um desses conjuntos de observações pode ser entendido como um perfil individual de resposta para a variável em estudo. O estudo desses perfis individuais pode ser um indício inicial para a inclusão de efeito aleatório no modelo (Rocha, 2010). Os dados longitudinais são tidos como regulares quando o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do tempo. E, se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais, tem-se uma estrutura balanceada em relação ao tempo. Dados do tipo longitudinais apresentam, conforme Singer et al. (2004), algumas vantagens, tais como:

- i. Permitem estudar o comportamento da variável resposta média da unidade experimental sobre cada tratamento;
- ii. Exige menos unidades experimentais (as medidas são feitas nas mesmas parcelas);
- iii. Permite verificar existência de dependência da variável resposta em relação às covariáveis existentes;
- iv. Diminuição do erro experimental e
- v. Melhora na precisão das estimativas dos parâmetros.

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

Segundo Cecon et al. (2008) para a utilização da análise em parcelas subdivididas no tempo assume-se que as respostas são igualmente correlacionadas em diferentes tempos. Já, para a análise de medidas repetidas não é necessário assumir essa pressuposição e, é possível considerar, por exemplo, que as respostas em tempos mais próximos, sejam, em geral, mais fortemente correlacionadas que as de períodos de tempos mais distantes. Uma consequência imediata de se ignorar diferentes correlações entre os dados medidos no tempo é que a significância aparente da diferença entre as médias dos tratamentos é extremamente exagerada e, também,

a sensibilidade dos testes da interação é seriamente reduzida, podendo estar até inválidos.

O objetivo deste trabalho foi analisar a produção de matéria seca de plantas, que foram submetidas a dois nutrientes que influenciam o crescimento, e que foram avaliadas ao longo de semanas, por meio da técnica estatística da análise de variância corrigida por duas formas distintas, e, simultaneamente, diferenciá-la da forma clássica de análise.

Metodologia

Os dados utilizados neste trabalho foram obtidos de um experimento conduzido em casa de vegetação do Departamento de Fisiologia Vegetal da Universidade Federal de Lavras na cidade de Lavras – MG, no período de março a abril de 2012. Foi considerado no estudo o efeito de um nutriente no crescimento de plantas, em dois níveis, caracterizando alta (A) e baixa (B) concentração, sendo a variável resposta peso seco, em gramas, avaliada por uma técnica não destrutiva, durante um período de cinco semanas com intervalos regulares de uma semana e, além disso, a cada tempo, um subconjunto diferente (10 plantas) é avaliado para cada tratamento. O modelo estatístico utilizado nesta análise foi o esquema de parcelas subdivididas, no delineamento blocos casualizados. Desta forma, foi utilizado o seguinte modelo:

$$y_{ijk} = \mu + \alpha_i + \beta_k + \delta_{ik} + \gamma_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

em que: y_{ijk} é a produção média da matéria seca, em gramas, da planta i , na semana j e bloco k ; μ é a constante inerente a qualquer parcela; α_i é o efeito do nutriente i ; β_k é o efeito do bloco k ; δ_{ik} representa o efeito aleatório da interação do k -ésimo bloco com o i -ésimo nutriente, tais que $\delta_{ik} \stackrel{\text{iid}}{\sim} N(0, \sigma_\delta^2)$; γ_k é o efeito do bloco k ; $(\alpha\beta)_{ij}$ é o efeito da interação do nutriente i com a semana j e ε_{ijk} é o erro aleatório, $\varepsilon_{ijk} \sim N(0, \Sigma)$.

Uma condição necessária para que o teste F da análise de variância seja exato é que a matriz de covariância Σ dos erros satisfaça a condição de esfericidade, caso não ocorra, há autores que propõem correção do número de graus de liberdade nas causas de variações que envolvem o fator tempo, ou seja, de todas as fontes de variação da sub parcela.

O teste de Mauchly está indicado em várias literaturas no que tange a analisar o critério de

esfericidade de Σ . Segundo (Freitas, 2008) o teste de Mauchly verifica se uma população normal multivariada apresenta variâncias iguais e se as correlações são nulas. Caso uma população satisfaça essa condição, será chamada de “esférica”. Esse teste utiliza a condição de esfericidade para Σ das medidas repetidas dos indivíduos requeridos nos $(t - 1)$ contrastes ortogonais normalizados. A estatística do teste de Mauchly é dada por:

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

em que S_{ij} é um elemento da i -ésima linha e j -ésima coluna da matriz de covariâncias amostral S para o erro intra indivíduos e escolhem-se $t - 1$ contrastes ortogonais normalizados nas t medidas repetidas. Seja a matriz:

$$C_{(t-1) \times t}$$

nos quais as linhas são contrastes ortogonais normalizados nas t medidas repetidas e obtém-se a matriz:

$$CSC'_{(t-1) \times (t-1)}$$

No intuito de melhorar a acurácia dessa aproximação pelo Teste de Qui-Quadrado a estatística do teste de Mauchly terá a seguinte hipótese nula:

$$H_0: CSC' = \gamma I$$

em que: $-\gamma \ln W$ terá distribuição χ_f^2 (Qui-quadrado, f graus de liberdade), com $f = \frac{1}{2}t(t - 1) - 1$. A hipótese nula será rejeitada ao nível α de significância se $-\gamma \ln W > \chi_{\alpha, f}^2$. Sendo que:

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

em que t = número de tempos; b = número de repetições; g = número de tratamentos.

Se a matriz de covariância não atender ao critério de esfericidade, são citados por vários autores, dentre eles (Malheiros, 2004) formas de corrigir os graus de liberdade das fontes de variações presentes na sub parcela da análise univariada de perfis no intuito de validar à aplicação do teste F . As correções de graus de liberdade devem ocorrer

somente a comparações “dentro”, ou seja, intra indivíduos (sub parcela). Já as comparações “entre” indivíduos (parcela) não são necessárias corrigir, pois o teste F aplicado a essas comparações possuem distribuição F exata. As formas de correção são:

i. Ajuste de Geisser e Greenhouse

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

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

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

ii. Ajuste de Huynh e Feldt

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

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.

Existe um terceiro ajuste citado por (Kirk, 1995) para quando a matriz de covariância possuir o formato esférico, ou seja, $\varepsilon = (t-1)^{-1}$. É demonstrável que $(t-1)^{-1} \leq \varepsilon \leq 1$, isto é, o valor máximo de 1 corresponde à condição de esfericidade e à medida que a matriz de covariância Σ se afasta do padrão, o valor de ε tende ao seu limite inferior. É demonstrável que se a condição de esfericidade for atingida, $\varepsilon = 1$. Entretanto, alguns autores relatam que quando as estimativas de ε estiverem próximas dos limites, isso induzirá a testes muito liberais, principalmente o estimador $\tilde{\varepsilon}$ comparado aos demais. Todavia, o estimador $\tilde{\varepsilon}$ é menos viesado que do que $\hat{\varepsilon}$, quando $\varepsilon \geq 0,75$. Nota-se também que $\tilde{\varepsilon}$ pode assumir valores maiores que 1, nesse caso deve adotá-lo igual a 1. Quando se acredita que o verdadeiro valor de $\varepsilon < 0,75$, ou não se tenha idéia sobre seu verdadeiro valor, há autores que indicam o uso do estimador $\tilde{\varepsilon}$, mas o

estimador $\hat{\varepsilon}$ possui aceitável controle do erro tipo I, embora possua a desvantagem de superestimar o real valor do nível de significância.

Resultados e discussões

Procedendo-se a análise do teste de Mauchly via software estatístico R (R Development Core Team, 2014), tabela 1, percebe-se que utilizando um nível de significância de 1% pode-se inferir a 1% de probabilidade que a matriz Σ não satisfaz a condição de esfericidade, ou seja, ela não possui variâncias homogêneas e covariâncias nulas, resultado coerente com a análise exploratória realizada, figura 1 e 2 abaixo, na qual possui evidências de heterogeneidade de variância. Pela figura 2, percebe-se também forte assimetria dos dados, algo que também infere que as variâncias não são homocedásticas.

Tabela 1. Resultado do teste de esfericidade de Mauchly

GL de f	Critério de Mauchly (W)	Teste de Qui-quadrado χ^2	Valor - p
104	1,12E-11	1239,1275	<0,0001

Ainda pela figura 1, percebe-se, em tese, que há indícios que o crescimento da matéria seca ocorre-se de forma quadrática na avolução das semanas avaliadas.

Pelos resultados é viável que se proceda a análise de variância usual; entretanto efetuando-se a correção dos graus de liberdade das fontes de variações contidas na sub parcela.

Pela análise usual de variância clássica (Tabela 2) verifica que não houve efeito da interação entre os nutrientes e as semanas avaliadas para a produção de matéria seca, produzida pelas plantas. Desta forma, deve-se analisar os efeitos principais dos nutrientes e as semanas, na qual consta-se que somente foi significativo o efeito do tempo. Registra-se que estes resultados são referentes a análise usual de variância; entretanto o resultado do teste de esfericidade recomendou corrigir os graus de liberdade das fontes de variações do teste F referente a subparcela.

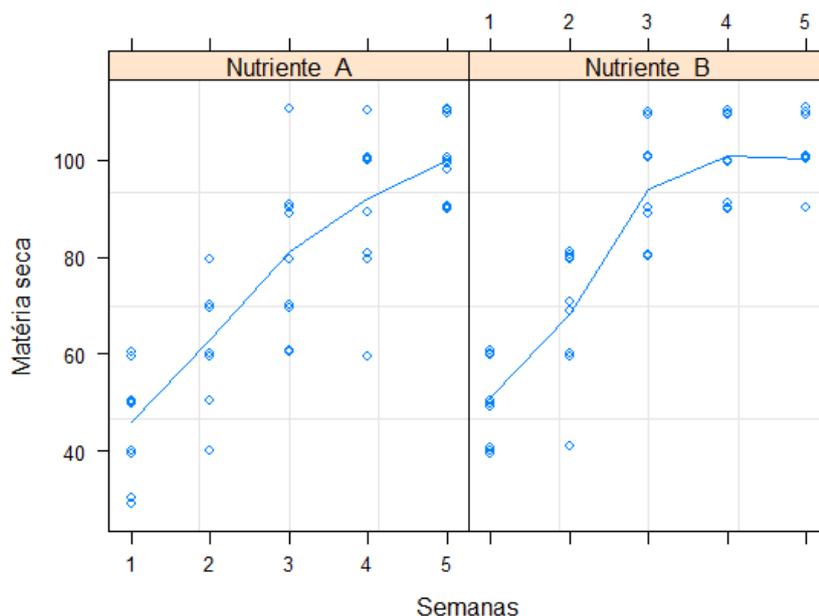


Figura 1. Produção de material seca pelos dois nutrientes nas cinco semanas avaliadas

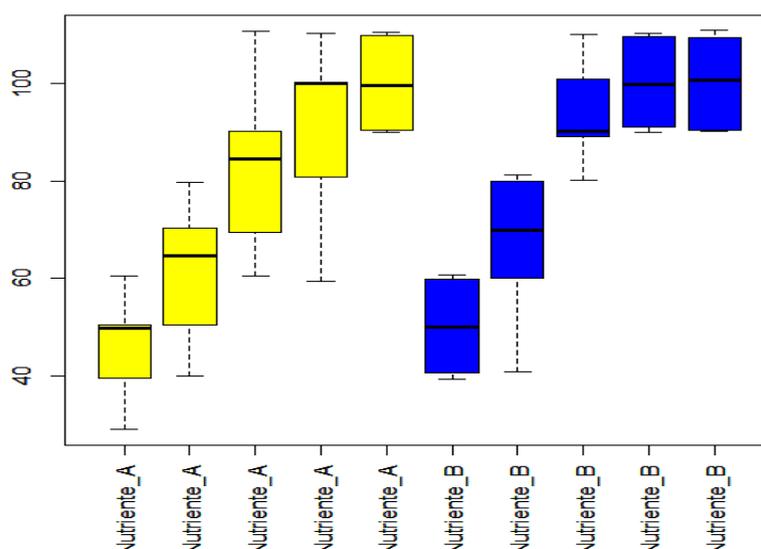


Figura 2. Boxplot da produção da matéria seca pelos dois nutrientes nas 5 semanas avaliadas.

Tabela 2. Resultado da análise de variância univariada da produção de matéria seca

Fontes de variação	GL	SQ	QM	F	Valor - p
Blocos	9	2626	291,8	0,96	0,52
Nutrientes	1	1061	1061,1	3,49	0,09
Erro (a)	9	2730	303,4		
Semanas	4	38684	9671,0	104,37	2e-16
Semanas*Nutrientes	4	439	109,8	1,18	0,32
Erro (b)	72	6671	92,7		
Semanas	4	38684	9671,0	104,37	2e-16

Tabela 3. Resultado da análise de variância univariada corrigida (subparcela) pelos ajustes de Geisser & Greenhouse (GG) e Huynh & Feldt (HF) da produção de matéria seca.

Teste para os fatores intra indivíduos (subparcela)					
Fontes de Variação	GL	SQ	QM	F	Valor - p
Semanas	4	38684	9671,0	104,37	2e-16
Semanas*Nutrientes	4	439	109,8	1,18	0,32
Erro (b)	72	6671	92,7		
Ajuste do teste F com correções dos graus de liberdade					
Fontes de variação	G-G	H-F			
Semanas	<0,001	<0,001			
Semanas*Nutrientes	0,245	0,203			

G-G= Grenhouse-Geisser $\hat{\epsilon} = 0,1265$ e H-F= Huynh – Feldt $\tilde{\epsilon} = 0,1426$

Pela [tabela 3](#) abaixo, percebe-se que a correção dos graus de liberdade não alterou os resultados obtidos pelas correções de Geisser & Greenhouse, bem como Huynh & Feldt, adotando um nível de significância de 1% de probabilidade

Desta forma, os resultados permanecem inalterados da análise de variância clássica realizada anteriormente. Sendo assim, deve-se estudar o efeito principal do fator semanas. Como o fator é quantitativo, deve-se usar o ajuste de

curvas por meio da análise de regressão em detrimento da aplicação de um teste de média convencional.

Pela [tabela 4](#) é realizado a decomposição da soma de quadrado do fator semanas para os possíveis modelos de regressão. Vê-se que o efeito cúbico e os desvios da regressão são não-significativos, desta forma o modelo que deve ser ajustado é o modelo quadrático, algo já evidenciado pela figura 1.

Tabela 4. Decomposição da soma de quadrados do fator semanas.

Fontes de variação	G	SQ	QM	F	Valor - p
Efeito Linear	1	35981	35981,9	388,3	<0,001
Efeito Quadrático	1	2298	2298,0	24,8	<0,001
Efeito cúbico	1	206	92,7	2,2	0,130
Desvio da regressão	1	197	197,7	2,1	0,140
Erro (b)	72	6671	92,6		

O modelo quadrático ajustado que representa a relação entre a produção média de matéria seca produzida pelas plantas utilizando os nutrientes A ou B e as 5 semanas avaliadas foi

$$y = 19,37 + 30.602081 x - 2,86 x^2$$

em que y representa a produção média, em gramas, de matéria seca e x representa as semanas avaliadas.

O modelo é reproduzido na figura abaixo onde se observa o crescimento de matéria seca, em gramas, produzidas por plantas no decorrer de 5 semanas avaliadas.

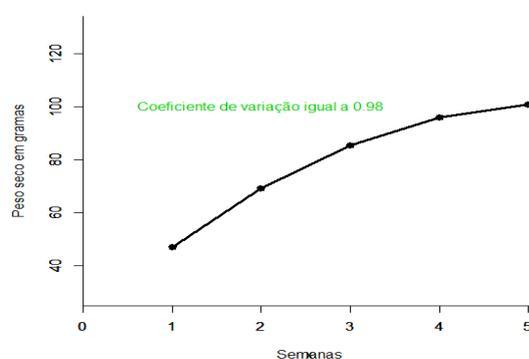


Figura 3. Descrição gráfica do crescimento de matéria seca, em gramas, produzidas por plantas no decorrer de 5 semanas avaliadas.

Conclusões

Há coerência de resultados entre a análise exploratória realizada e a aplicação do teste de Mauchly, no que tange a heterogeneidade de variâncias, principalmente em semanas mais distantes. A produção de matéria seca tende a se estabilizar após o período avaliado. A correção dos graus de liberdade das fontes de variações da sub parcela na análise de variância não diferiu significativamente dos resultados da análise clássica.

O crescimento de matéria seca produzido não é afetado, em média, pelo uso dos dois nutrientes de alta (A) e baixa (B) concentração.

Referências Bibliográficas

- Cecon, P.R., Silva, F.F., Ferreira, A., Ferrão, G.R., Carneiro, A.P.S., Detmann, E., Faria, P.N. & Morais, T.S.S. (2008). Análise de medidas repetidas na avaliação de clones de café 'Conilon'. *Pesquisa Agropecuária Brasileira*, 43, 1171-1176.
- Freitas, E.G. (2008). Análise de dados longitudinais em experimentos com cana-de-açúcar. Dissertação de Mestrado: Escola Superior de Agricultura Luiz de Queiroz – Piracicaba.
- Faraway, J. J. (2006). Extending the linear model with R: generalized linear, mixed effect and nonparametric regression models. New York: Chapman & Hall/CRC, 331 p.
- Malheiros, E.B. (2004). 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*, 22, 23-29.
- Kirk, R.E. (1995). *Experimental Design: Procedures For The Behavioral Sciences* (3 ed.). Pacific Grove, CA, USA: Brooks/Cole.
- Nobre, J. S. & Singer, J. M. (2007). Residuals analysis for linear mixed models. *Biometrical Journal*, 49, 863-875.
- R Development Core Team: a language and environment for statistical computing. R Foundation for Statistical Computing, versão 2.13.1. Vienna, Austria, 2014 Disponível em: <<http://www.R-project.org>>.
- Rocha, E. B. (2010). Modelos não-lineares para dados longitudinais provenientes de experimentos em blocos casualizados abordagem bayesiana. 2010. 96 p. Dissertação (Mestrado em Ciências) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba.
- Singer, J. M., Rocha, F. M. M. & Nobre, J. S. (2004). Análise de medidas repetidas. In: *Jornada de Estatística*, 4, Maringá. Anais... Maringá: UEM, p. 120.

Article History:

Received 13 May 2016

Accepted 20 June 2016

Available on line 7 July 2016

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.