

**Análises estatísticas de dados de cortes sucessivos de alfafa: modelo linear padrão ou modelo linear misto?**Alfredo Ribeiro de Freitas¹, Reinaldo de Paula Ferreira², Adônis Moreira¹, Armando de Andrade Rodrigues²

¹Embrapa Pecuária Sudeste, CEP: 13560-970, São Carlos, SP, *Bolsista do CNPq, emails: ribeiro@cnpse.embrapa.br; adonis@cnpse.embrapa.br.

²Embrapa Pecuária Sudeste, CEP: 13560-970, São Carlos, SP; emails: reinaldo@cnpse.embrapa.br; armando@cnpse.embrapa.br

Resumo: Dados de produção de matéria seca (PMS) de 20 Cortes mensais de 92 cultivares de alfafa (*Medicago sativa* L.), obtidos de um experimento em blocos casualizados realizado na Embrapa Pecuária Sudeste, São Carlos, SP, de 2004 a 2006, foram analisados considerando-se Cortes como medidas repetidas. O objetivo foi escolher a estrutura mais adequada da matriz de variância e covariância (R) dos erros dentro de Cortes e decidir o procedimento mais adequado para análise: GLM ou MIXED do SAS. A estrutura de R escolhida foi a Autoregressiva de Primeira Ordem de Média Móvel - ARMA(1,1), com três parâmetros estimados: autoregressivo ($\rho = 0,9402$), variância residual ($\sigma^2 = 76385$) e componente de média móvel ($\gamma = 0,6642$). Considerando-se essa estrutura obteve significância para o efeito de cultivares ($p = 0,0151$); Corte ($p < 0,0001$) e interação cultivares x corte ($p < 0,0003$); a segunda opção foi a Autoregressiva de Primeira Ordem - AR(1) com parâmetros ρ e σ^2 . Conclui-se que o procedimento MIXED é mais eficiente que o GLM para analisar dados de PMS obtidos de cortes mensais de alfafa.

Palavras-chave: matriz ARMA(1,1), procedimento GLM, procedimento MIXED, produção de matéria seca

Statistical analysis of successive cuts of alfalfa data: standard linear model or mixed linear model?

Abstract: Data of dry matter production (DMP) obtained from 20 monthly Cuts of 92 accessions of alfalfa (*Medicago sativa* L.), obtained of a randomized block experiment carried out at Embrapa Cattle Southeast, in São Carlos, SP, from 2004 to 2006, were analyzed considering Cuts as repeated measures. The objective was to select the most adequate structure of variance and covariance (R) of error within Cuts and to decide which will be the correct procedure for the analysis: GLM or MIXED of SAS. The selected R structure was the Autoregressive Moving Average - ARMA(1,1) with three estimated parameters: autoregressive ($\rho = 0.9402$), residual variance ($\sigma^2 = 76385$) and a moving average component ($\gamma = 0.6642$). Considering this structure all effects was significant: accessions ($p = 0.0151$); cuts ($p < 0.0001$) and accessions x cuts interactions ($p < 0.0003$); the second alternative selected was the Autoregressive first Order - AR(1,1) with parameters ρ and σ^2 . It was concluded that the MIXED procedure it is more efficient than GLM for analysing data of DMP of alfalfa obtained from monthly cuts.

Keywords: ARMA(1,1) matrix, dry matter production, GLM procedure, MIXED procedure,

Introdução

Na área de forragicultura é comum o corte sucessivos de forragens para o estudo de características como a produção de matéria seca, teor de nitrato, teor de proteína bruta, persistência, fibra em detergente neutro, etc. Geralmente, o delineamento utilizado é o de blocos casualizados, em que os tratamentos ou combinações destes são alocados aleatoriamente às parcelas e as avaliações no tempo (cortes) são as medidas repetidas. Nesta análise, cada combinação bloco-tratamento representa o indivíduo ou a parcela e dois tipos de variabilidade são considerados: entre e dentro de parcela. O modelo misto usual (Littell et al., 1996; Freitas et al. 2007) é dado por: $y_{ijk} = \mu + \alpha_i + \delta_{ij} + t_k + (\delta t)_{ik} + \varepsilon_{ijk}$; y_{ijk} é a resposta no corte k na parcela j onde foi aplicado o tratamento i; μ é a média global; α_i é o efeito do tratamento i; δ_{ij} é o efeito aleatório da parcela j dentro do tratamento i; t_k é o efeito fixo do corte k; $(\delta t)_{ik}$ é o efeito de interação e ε_{ijk} é o erro aleatório entre as medidas ou erro dentro de parcela. Quando a distribuição dos dados é pelo menos aproximadamente normal, o procedimento MIXED do SAS (SAS 2002–2003) disponibiliza cerca de 40 tipos de matrizes de estruturas de variâncias e covariâncias do tipo $\text{Var}(\varepsilon_{ijk}) = R$ e fornece três critérios para selecionar a mais adequada (Bozdogan, 1987). Dependendo da estrutura de R, os dados podem ser analisados por meio do modelo linear padrão, modelo linear misto, ou ambos; para os usuários do SAS, significa utilizar os procedimentos GLM e MIXED, respectivamente.



Quando R é do tipo Componente de Variância (VC), a correlação entre as medidas repetidas é zero e os dados podem ser analisados por meio do modelo linear padrão; se R é do tipo *Huynh-Feldt* (HF) e Simetria Composta (CS), significa que a condição de esfericidade é atendida; nesse caso os dados podem ser analisados por ambos os procedimentos: GLM e MIXED do SAS. Se a matriz R escolhida for uma estrutura diferente de VC, HF e CS, o procedimento correto para análise é o MIXED. Biologicamente, pode-se admitir independência entre cortes, uma vez que toda a forragem da parcela é cortada. No entanto, ao longo dos cortes, os efeitos de nutrientes no solo e o efeito de sazonalidade na produção, principalmente, devido à existência de períodos de seca (abril a setembro) e de águas (outubro a março), podem proporcionar correlação entre os efeitos de cortes.

O objetivo foi analisar dados da produção de matéria seca (PMS) de 20 cortes de 92 cultivares de alfafa (*Medicago sativa* L.) como medidas repetidas, escolher a estrutura mais adequada da matriz de variância e covariância (R) e, finalmente decidir o procedimento mais adequado para análise: GLM ou MIXED.

Material e Métodos

Foram analisados dados de produção de matéria seca (PMS), kg/ha, de alfafa colhidos de um experimento com 92 genótipos de alfafa conduzido em blocos casualizados, duas repetições, realizado no ano de 2006 na Embrapa Pecuária Sudeste, São Carlos, SP, com 20 cortes mensais consecutivos, realizados quando aproximadamente 10% das plantas estavam em florescimento. Os dados foram analisados por meio do procedimento *MIXED* do SAS considerando as avaliações no tempo (cortes) como medidas repetidas e o modelo misto usual de Littell et al., (1996), citado no item anterior. Foram avaliadas as seguintes estruturas de R: VC, HF, Não-Estruturada (UN), Autoregressiva de Primeira Ordem - *AR(1)* e Autoregressiva de Primeira Ordem de Média Móvel - *ARMA(1,1)*; a estrutura de R mais adequada foi selecionada pelo critério de informação de Akaike - AIC e teste de razão de verossimilhança, descritos em Bozdogan (1987).

Resultados e Discussão

Na representação por meio de diagrama de caixa (Figura 1), para cada corte, a parte inferior e a parte superior da caixa corresponde, respectivamente, a 25% (1º quartil) e 75% (3º quartil) dos dados ordenados, verifica-se que a mediana (linha horizontal no meio da caixa) de PMS ao longo dos cortes, segue uma tendência que pode ser explicada, principalmente, pelo efeito de sazonalidade na produção, devido à existência de períodos de seca (abril a setembro) e de águas (outubro a março), que ocorreram durante a realização dos 20 cortes.

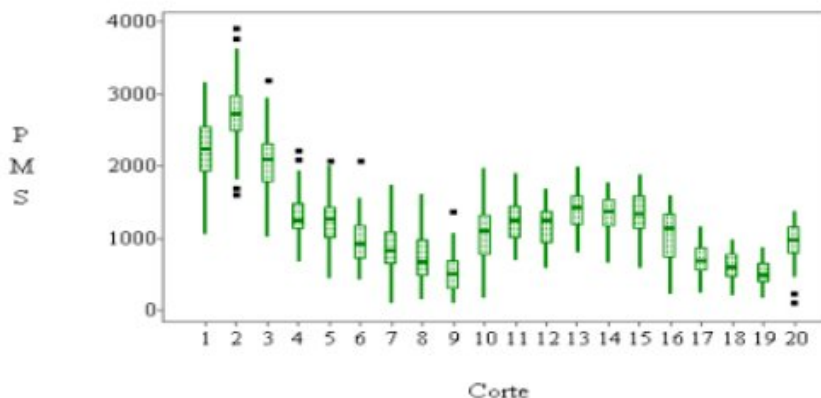


Figura 1 Representação por meio de diagrama de caixa de dados de produção de matéria seca (PMS), kg/ha, de 20 cortes de 92 cultivares de alfafa (*Medicago sativa* L.)

Na Tabela 1 é apresentado o número de parâmetros, valor da função de máxima verossimilhança ($-2\text{Log} L$) e o critério de informação de Akaike (AIC) para seis estruturas de covariâncias ajustadas. Verifica-se que a *ARMA(1,1)*, a *AR(1)* e a CS, pela ordem, são as que apresentam o melhor ajuste aos dados, pois são as que possuem os menores valores de AIC. A escolha dessa matriz também foi confirmada quando construiu o teste de razão de verossimilhança restrito que permite comparar as matrizes duas a duas por meio da estatística, ($-2\text{Log} L$: matriz 1) - ($-2\text{Log} L$: matriz 2); este teste equivale a testar a hipótese H_0 : matriz 1=matriz 2 versus H_a : matriz 1 \neq matriz 2, em que o resultado é um teste de Qui-quadrado (χ^2), com graus de liberdade igual à diferença do número de parâmetros das duas matrizes.



Na comparação de ARMA(1,1) versus AR(1), CS, UN,VC e HF, obteve-se, respectivamente, $\chi^2_1 = 108,4$; $\chi^2_1 = 98,9$; $\chi^2_{207} = 568,1$; $\chi^2_1 = 440,5$; $\chi^2_{19} = 795,8$, valores altamente significativos ($P < 0,001$), indicando que a estrutura ARMA(1,1) é significativamente superior às demais para modelar erros dentro de parcela na produção de matéria seca de cortes sucessivos de alfafa. Quando se observa a Figura 1, verifica-se que estes resultados são esperados, pois matrizes como a $AR(1)$ que contém um parâmetro autoregressivo (ρ) e a variância residual σ^2 e a $ARMA(1,1)$, que além de ρ e σ^2 , inclui ainda um parâmetro que modela o componente de média móvel (γ), são fortes candidatas para modelar erros entre cortes sucessivos de forrageiras quando se tem tendência sazonal na produção. Considerando-se essa estrutura obteve significância para o efeito de tratamento ($p=0,0151$); Corte ($p<0,0001$) e interação tratamento x corte ($p<0,0003$).

Tabela 1 Número de parâmetros, valor da função de máxima verossimilhança ($-2\text{Log } L$) e critério de informação de Akaike (AIC)

Estrutura	Número de parâmetros	$-2 \text{Log } L$	AIC
<i>ARMA(1,1)</i>	2	13957,5	13963,5
<i>AR(1)</i>	1	14123,8	14127,8
CS	1	14128,1	14132,1
HF	20	14416,4	14458,4
UN	209	14110,3	14850,3

Conclusões

Na análise de dados de produção de matéria seca (PMS) de cortes mensais de alfafa (*Medicago sativa* L.), a estrutura de variância e covariância mais adequada para modelar erros entre cortes é a ARMA(1,1); a segunda opção foi a Autoregressiva de Primeira Ordem - AR(1). Dessa forma, o procedimento MIXED é mais eficiente que o GLM para analisar dados de PMS.

Literatura citada

- BOZDOGAN, H. 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. **Psychometrika**, v.52, n.3, p.345-370, 1987
- FREITAS, A. R.; BARIONI JUNIOR, W.; FERREIRA, R.P.; DESTEFANI, C.; SANTOS, A.R.; MOREIRA, A. Estudo de medidas repetidas: uma aplicação a dados de forrageiras. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DASOCIEDADE INTERNACIONAL DE BIOMETRIA, 52.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12, 2007, Santa Maria, RS. **Resumos...** Santa Maria: UFSM; RBRAS, 2007.
- LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS System for Mixed Models**. Cary: Statistical Analysis System Institute, 1996. 633p.
- SAS Institute. User's Guide. versão 9.1.3, versão para Windows. Cary, NC, USA, 2002–2003.