

## Modelos lineares mistos para análise de dados longitudinais provenientes de ensaios agronômicos

Ezequiel Abraham López Bautista –FAUSAC, LCE-ESALQ-USP<sup>1,2</sup>

Edilan de Sant'Ana Quaresma –ICED/UFOPA, LCE-ESALQ-USP

Mauricio Lordêlo Santana –UEFS, LCE-ESALQ-USP

### 1 Introdução

Em estudos longitudinais, repetidas observações de uma mesma variável resposta são coletadas na mesma unidade experimental, em diferentes ocasiões. Como diferentes observações são realizadas na mesma unidade, espera-se que estas sejam correlacionadas, e que exista uma heterogeneidade de variâncias nas diferentes ocasiões.

Para analisar este tipo de dados, dada a existência de correlação entre diferentes observações feitas em uma mesma unidade experimental, o uso do modelo linear clássico resulta não ser adequado, já que pressupõe independência entre as medidas repetidas. Além disso, é comum encontrar heterogeneidade de variâncias nas diversas ocasiões em que são feitas as medidas, fazendo com que a pressuposição de homocedasticidade do modelo linear clássico também não seja satisfeita. O modelo linear misto (ou modelo de efeitos aleatórios) é uma opção na modelagem deste tipo de dados, por possibilitar a utilização de diferentes estruturas para matrizes de covariâncias, que permite modelar tanto a correlação entre as medidas repetidas quanto a heterocedasticidade.

O presente trabalho tem como objetivo apresentar a aplicação da metodologia de modelos lineares mistos para analisar os dados de porcentagem de cobertura estimada por parcela, da espécie forrageira *Brachiaria decumbens* cv. Basilik, obtidos em cinco leituras num experimento com delineamento casualizado em blocos, para comparar cinco métodos de labrança.

Para analisar estes dados, primeiro foram ajustados modelos com distintas estruturas de covariância, combinando apropriadamente estruturas de correlação residual, heterocedasticidade residual e efeitos aleatórios. Utilizando critérios de verossimilhança penalizada (AIC e BIC) foi selecionado o modelo que melhor descreveu os dados, e usando este modelo foram realizadas inferências sobre as médias (comparação de tratamentos, estudar o efeito do tempo, analisar se os perfis médios variavam no tempo, etc.).

---

<sup>1</sup> Contato: ealbautis@usp.br

<sup>2</sup> Agradecimento ao PEC-PG da CAPES pela bolsa concedida.

## 2 Material e Métodos

**Tratamentos:** os dados são provenientes de um ensaio de estabelecimento de forrageiras, reportado por Moser e Macchiavelli (2002), para comparar cinco métodos de labrança ( $T_1$  = labrança mínima;  $T_2$  = labrança mínima com herbicida;  $T_3$  = labrança mínima com herbicida e arado de disco aos 45 dias;  $T_4$  = labrança zero, e  $T_5$  = labrança convencional) na região central úmida de Porto Rico.

**Delineamento experimental:** casualizado em blocos com três repetições por tratamento.

**Variável resposta:** porcentagem de cobertura estimada por parcela, da espécie forrageira *Brachiaria decumbens* cv. Basilik, obtidas em cinco leituras, tomadas com intervalos de um mês entre agosto e dezembro de 2001.

**Análise estatística:** inicialmente foram construídos os perfis médios de cobertura observados nos cinco tempos para cada um dos tratamentos. Como estratégia geral para analisar estes dados, primeiro foram ajustados modelos com distintas estruturas de covariância, combinando apropriadamente estruturas de correlação residual, heterocedasticidade residual e efeitos aleatórios. Por meio de critérios de verossimilhança penalizada (AIC e BIC) e utilizando o método da máxima verossimilhança restrita (REML), foi selecionado o modelo que melhor descreveu os dados, e usando este modelo foram realizadas inferências sobre as médias (comparação de tratamentos, estudar o efeito do tempo, analisar se os perfis médios variavam no tempo, etc.). Para escolher o melhor modelo, inicialmente foi proposto um modelo simples com poucos parâmetros a estimar (i.e. parcimonioso), e posteriormente adicionando parâmetros até atingir o modelo sem estrutura, que é o menos parcimonioso. As seguintes estruturas de covariância para os dados (covariância marginal) foram utilizadas:

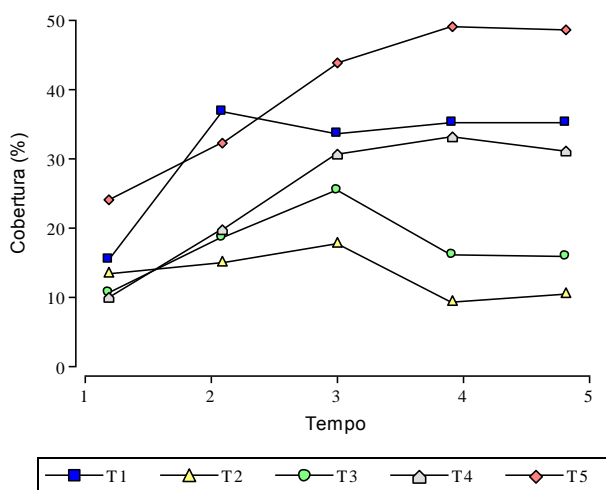
1. Efeito aleatório de bloco e erros independentes e homocedásticos.
2. Efeitos aleatórios de bloco e parcela dentro de cada bloco, erros independentes e homocedásticos.
3. Efeito aleatório de bloco e erros independentes e heterocedásticos.
4. Efeitos aleatórios de bloco e parcela dentro de cada bloco, erros independentes e heterocedásticos.
5. Efeitos aleatórios de bloco, correlação constante entre erros da mesma parcela e variância residual constante no tempo.
6. Efeitos aleatórios de bloco, correlação constante entre erros da mesma parcela e variância residual diferente nos distintos tempos.

7. Efeitos aleatórios de bloco, estrutura autorregressiva de primeira ordem entre os erros da mesma parcela e variância residual constante no tempo.
8. Efeitos aleatórios de bloco e parcela dentro de cada bloco, estrutura autorregressiva de ordem 1 entre os erros da mesma parcela e variância residual constante no tempo.
9. Efeitos aleatórios de bloco, estrutura autorregressiva de ordem 1 entre os erros da mesma parcela e variância residual diferente no distintos tempos.
10. Efeitos aleatórios de bloco e parcela dentro de cada bloco, estrutura autorregressiva de ordem 1 entre os erros da mesma parcela e variância residual diferente nos distintos tempos.
11. Efeitos aleatórios de bloco, sem estrutura para as correlações entre erros provenientes da mesma parcela e variâncias residuais diferentes no tempo.

Após selecionar a estrutura de covariância dos dados foram realizadas inferências sobre as médias, utilizando o teste de F-Wald e posteriormente o teste de comparações múltiplas LSD (*Least Significant Difference*) proposto por Fisher, com  $\alpha = 0,05$ .

### 3 Resultados e discussão

Os perfis médios de cobertura observados nos cinco tempos para cada um dos tratamentos é apresentado na Figura 1.



**Figura 1** – Relação entre cobertura e tempo para cinco tratamentos (métodos de labrança)

Na Figura 1 pode-se observar um comportamento diferencial ao longo do tempo, para cada tipo de método de labrança. A seguir, na Tabela 1 são apresentados os critérios para seleção do melhor modelo.

**Tabela 1** – Critérios de informação de Akaike (AIC) e Bayesiano (BIC) e logaritmo da função de verossimilhança (LV) pelo método REML dos modelos ajustados.

Modelo	AIC	BIC	LV
1	476,39	528,01	-211,19
2	470,00	523,54	-207,00
3	475,13	534,40	-206,57
4	470,54	531,73	-203,27
5	470,00	523,54	-207,00
6	466,67	527,85	-201,33
7	460,93	514,47	-202,47
8	462,93	518,38	-202,47
9	492,70	523,89	-199,35
10	464,70	527,80	-199,35
<b>11</b>	<b>434,74</b>	<b>513,14</b>	<b>-176,37</b>

Ao comparar os valores de AIC (ou de BIC) para as estruturas ajustadas, observa-se que o menor valor é obtido com o modelo 11 (AIC = 434.74, BIC = 513.14), por tanto selecionamos o modelo de covariância sem estrutura. Este modelo proporcionou as seguintes variâncias residuais para cada um dos 5 tempos avaliados:  $\sigma_1^2 = 42,0072$ ,  $\sigma_2^2 = 64,5903$ ,  $\sigma_3^2 = 172,7327$ ,  $\sigma_4^2 = 278,7291$  e  $\sigma_5^2 = 254,5005$ . Nota-se a variabilidade entre as leituras.

Na Tabela 2 se apresenta o resumo dos resultados do teste de hipóteses sequenciais para os efeitos fixos, na qual se observa que a interação entre tratamentos e tempo não foi significativa, portanto foram aplicados posteriormente testes de comparações múltiplas de médias para os efeitos de tratamento e tempo, por separado.

**Tabela 2** – Teste de hipótese sequencial para os efeitos fixos.

Causas de variação	g.l.	F	Valor p
Tratamento	4	4,05	0,0065
Tempo	4	16,77	< 0,0001
Tratamento × Tempo	16	1,49	0,1417

Nas Tabelas 3 e 4 são apresentados os resumos do teste de comparações múltiplas LSD, para o tempo e tratamentos, respectivamente. A partir desses resultados, tem-se que, em média para os quatro tratamentos, se observa uma mudança significativa entre os tempos 1 e 2, porém nos tempos posteriores a cobertura média não muda significativamente (Tabela 3). Por outra parte, na Tabela 4, observa-se que o T<sub>5</sub> (labrança convencional) forneceu os melhores resultados de cobertura da espécie forrageira *B. decumbens*, e o T<sub>2</sub> (labrança mínima com herbicida) os piores resultados.

**Tabela 3** – Cobertura (%) média obtida ao longo do tempo e resumo do teste de LSD.

Tempo	Cobertura (%) média	Erro padrão	Grupos
3	30,29	3,70	a
4	28,53	4,56	a
5	28,27	4,38	a
2	24,53	2,55	a
1	14,73	2,33	b

Médias com uma letra comum não são significativamente diferentes ( $p \leq 0,05$ )

**Tabela 4** – Cobertura (%) média obtida com os tratamentos aplicados e resumo do teste de LSD.

Tratamento	Cobertura (%) média	Erro padrão	Grupos
T <sub>5</sub>	39,60	5,47	a
T <sub>1</sub>	31,27	5,47	a b
T <sub>4</sub>	24,96	5,47	a b c
T <sub>3</sub>	17,33	5,47	a b c
T <sub>2</sub>	13,19	5,47	c

Médias com uma letra comum não são significativamente diferentes ( $p \leq 0,05$ )

#### 4 Conclusões

O estudo mostrou que o modelo de efeitos aleatórios de bloco, sem estrutura para as correlações entre erros provenientes da mesma parcela e variâncias residuais diferentes no tempo, foi o que melhor se ajustou para descrever os dados de porcentagem de cobertura da espécie forrageira *Brachiaria decumbens* cv. Basilik, e diferenças significativas foram observadas unicamente para os efeitos principais (tratamentos e tempo).

#### 5 Referências

- [1] GBUR, E.; STROUP, W.; McCARTER, K.; DURHAM, L.; YOUNG, M. 2013. **Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences. Madison.** Madison: American Society of Agronomy, Soil Science Society of America and Crop Science Society of America. 283 p.
- [2] MOSER, E.B.; MACCHIAVELLI, R. 2002. Model selection techniques for repeated measures covariance structures. **Proceedings of the XIV Conference on Applied Statistics in Agriculture**, v.14, p.17-31.
- [3] SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide: **The GLIMMIX Procedure (Chapter)**. Cary, NC: SAS Institute Inc.