

## REGRESSÃO ALEATÓRIA E FUNÇÕES DE COVARIÂNCIA NA ANÁLISE DE MEDIDAS REPETIDAS

Marcos Deon Vilela de RESENDE<sup>1</sup>  
Gabriel Dehon S. P. REZENDE<sup>2</sup>  
José Sebastião Cunha FERNANDES<sup>3</sup>

- **RESUMO:** A teoria da regressão aleatória preconiza que existe, em um dado experimento, diferentes coeficientes de regressão associados aos diferentes indivíduos da população, de forma que se os indivíduos são efeitos aleatórios, os coeficientes de regressão associados a estes indivíduos, devem também ser tratados como efeitos aleatórios. Este trabalho teve como objetivos apresentar a técnica da regressão aleatória e estudá-la para fins de aplicação no melhoramento genético de espécies perenes. Empregando dados simulados e experimentais, a referida técnica permitiu: descrever continuamente a estrutura de covariâncias de caracteres de crescimento ao longo do tempo; a estimação simultânea de parâmetros genéticos ao longo das idades, considerando a mudança genética do caráter ao longo do tempo; a predição de valores genéticos de indivíduos avaliados em diferentes idades e a projeção destes valores genéticos para uma idade comum para efeitos de ordenamento e seleção; a obtenção de curvas de valores genéticos para cada indivíduo em função da idade. O modelo de regressão aleatória empregado conduziu a melhores resultados que os tradicionais modelos univariados e de repetibilidade e permitiu a obtenção de estimativas dos mesmos parâmetros que podem ser obtidos por um modelo multivariado, porém, com uma menor parametrização, fato que é vantajoso computacionalmente.
- **PALAVRAS-CHAVE:** Modelos mistos; modelos multivariados; predição linear; medidas repetidas; dados longitudinais; seleção genética.

---

1 EMBRAPA, Caixa Postal 319, CEP 83411-000 – Colombo – PR.

2 Aracruz Celulose S.A., Aracruz – ES.

3 Departamento de Genética – Universidade Federal do Paraná, Caixa Postal 19071, CEP 81531-990 – Curitiba – PR.

## 1 Introdução

Nos modelos gerais de regressão (simples ou múltipla), as covariáveis ou variáveis independentes na realidade não são variáveis, mas sim constantes, ou seja, em hipotética amostragem repetida os valores dessas covariáveis permanecem constantes. Nestes modelos, os coeficientes de regressão são fixos, ou seja, os mesmos coeficientes de regressão estão associados a todos os indivíduos da população. Por outro lado, existem modelos nos quais a suposição adequada referente aos coeficientes de regressão associados às covariáveis é que cada membro possui distribuição aleatória. Estes modelos são denominados modelos de regressão aleatória.

A teoria associada à regressão aleatória foi apresentada inicialmente por Henderson Junior (1982). Baseado nesta teoria existe, em um dado experimento, diferentes coeficientes de regressão associados aos diferentes indivíduos da população, de forma que se os indivíduos são efeitos aleatórios, os coeficientes de regressão associados a estes indivíduos, devem também ser tratados como efeitos aleatórios (Henderson, 1984).

Como é o caso no melhoramento genético de espécies perenes, os indivíduos são tratados como efeitos aleatórios através de seus valores genéticos. Também de maneira geral, avaliações repetidas são realizadas em cada indivíduo no decorrer do tempo ou idade, de forma que variáveis respostas podem ser explicadas através de uma função linear da idade. Neste contexto, os modelos de regressão aleatória são capazes de explicar melhor a variabilidade associada a medidas repetidas do que os modelos lineares simples (que incluem o efeito de ambiente permanente e o parâmetro repetibilidade ou tratam as medidas repetidas como caracteres diferentes através de um modelo multivariado) empregados em avaliação genética. Os modelos de regressão aleatória permitem a obtenção de diferentes curvas de valores genéticos associadas aos diferentes indivíduos e considera as mudanças nas variâncias genéticas e residuais através do tempo. Também, desde que existam confiáveis estimativas de componentes de variância, estes modelos permitem a predição de valores genéticos de um indivíduo nas diferentes idades, baseando-se em avaliações de apenas uma idade.

No melhoramento genético de espécies perenes, a regressão aleatória no contexto dos modelos lineares mistos apresenta grande utilidade, pois permite a predição de valores genéticos de indivíduos avaliados em diferentes idades (e com diferentes números de idades

avaliadas) e a projeção destes valores genéticos para uma idade comum, para efeito de ordenamento e seleção. Apesar da aplicação dos modelos de regressão aleatória ser comum em outras áreas (Longford, 1993), a aplicação destes modelos ao melhoramento genético iniciou-se recentemente com os trabalhos de Schaeffer & Dekkers (1994), Jamrozick & Schaeffer (1997) e Jamrozick et al. (1997), visando a avaliação genética em gado bovino leiteiro.

A estimação de covariâncias entre os coeficientes de regressão aleatória produz estimativas de funções de covariâncias (Kirkpatrick et al., 1990), que referem-se a uma descrição contínua da estrutura de covariâncias do caráter para a amplitude de idades abrangida pelos dados, ou seja, uma função de covariância descreve a covariância entre medições tomadas em certas idades como função destas idades. Meyer (1998a) demonstrou a equivalência entre os modelos de regressão aleatória e de funções de covariância.

Os modelos de regressão aleatória e as funções de covariância podem expressar, de maneira mais realística, os fenômenos associados a dados longitudinais (por exemplo, curvas de crescimento através do tempo ou medições repetidas). Meyer (1998a) relata que estes modelos deverão suplantiar os modelos menos realísticos até então usados. Os modelos de regressão aleatória permitem considerar que o caráter em questão pode estar mudando continuamente e gradualmente através do tempo (Kirkpatrick et al., 1994), e portanto podem ser mais realísticos que os modelos de repetibilidade e multivariado.

O presente trabalho tem como objetivos apresentar os aspectos metodológicos associados aos modelos de regressão aleatória e estudar as possibilidades de aplicação ao melhoramento genético de espécies perenes.

## 2 Metodologia

### 2.1 Regressão aleatória no Contexto dos Modelos Lineares Mistos

Para caracteres avaliados em vários estágios de desenvolvimento dos indivíduos, o seguinte modelo linear pode ser adotado (para o caso univariado):

$$Y_{ijk} = P_i + (b_0 + b_1 A^*) + (a + \lambda A^*)_j + e_{ijk}$$

onde:

- $Y_{ijk}$  é a k-ésima medição do indivíduo  $j$  associado ao efeito fixo  $i$  ( $i = 1, 2, 3, \dots, b$ ;  $j = 1, 2, 3, \dots, n$ ;  $k = 1, 2, 3, \dots, m$ ).
- $P_i$  representa os efeitos fixos do modelo.
- $b_0$  e  $b_1$  são coeficientes de regressão fixa da medição na idade.
- $A^*$  é a idade em meses ou anos.
- $a$  e  $\lambda$  são coeficientes de regressão aleatória relacionando a medição do indivíduo  $j$  a um intercepto e idade, respectivamente. Em conjunto são os efeitos aleatórios de valor genético aditivo.  $a$  representa diferenças genéticas entre indivíduos não explicadas pela idade e  $\lambda$  representa diferenças genéticas entre indivíduos em função da idade.
- $e_{ijk}$  é o erro aleatório.

Através deste modelo, a curva de valores genéticos para determinado indivíduo pode ser vista através de dois grupos de regressão: regressão fixa para todos os indivíduos pertencendo a um mesmo nível de efeito fixo, a qual descreve a forma geral para um determinado indivíduo; regressões aleatórias descrevendo os desvios da regressão fixa, gerando diferentes curvas para os diferentes indivíduos.

Em notação matricial, o modelo pode ser escrito:

$$y = X\beta + Z_0 a + Z_1 \lambda + e, \text{ onde:}$$

- $y, \beta, e$  vetores de dados, de efeitos fixos e de erros aleatórios, respectivamente.
- $a, \lambda$  vetores referentes aos coeficientes de regressão aleatória, que relacionam as medições aos interceptos e idades, respectivamente.
- $X$  é a matriz de incidência para os efeitos fixos.
- $Z_0$  é a matriz de incidência para  $a$ , contendo 0 e 1's.
- $Z_1$  é uma matriz associando  $\lambda$  a  $y$ , contendo zero e valores de idade.

Nas matrizes  $Z_0$  e  $Z_1$  as primeiras colunas são nulas devidas ao fato de que referem-se aos indivíduos que não possuem observação.

Tem-se as seguintes médias e estrutura de variâncias associadas ao modelo:

$$E(y) = X\beta \text{ e } E(Y_{ijk}) = P_i + (b_0 + b_1 A^*),$$

em que  $(b_0 + b_1 A^*)$  refere-se à média geral na idade  $A^*$ .

$$\text{Var} \begin{pmatrix} a \\ \lambda \\ e \end{pmatrix} = \begin{bmatrix} Ag_{aa} & Ag_{a\lambda} & 0 \\ Ag_{a\lambda} & Ag_{\lambda\lambda} & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}, \text{ onde:}$$

$A$  matriz de parentesco genético aditivo  
 $g_{ij}$  são covariâncias entre os coeficientes de regressão aleatória, ou seja, são funções de covariância que descrevem continuamente a estrutura de covariâncias para o caráter, no intervalo de idades abrangido pelos dados.

A matriz de covariância ( $G$ ) entre os efeitos genéticos aleatórios, desconsiderando as relações de parentesco equívale a:

$$G = \begin{pmatrix} g_{aa} & g_{a\lambda} \\ g_{a\lambda} & g_{\lambda\lambda} \end{pmatrix}$$

## 2.2 Estimação dos Efeitos Fixos e Predição dos Efeitos Aleatórios

A estimação e predição dos efeitos segundo o modelo linear definido é realizado através das equações de modelo misto:

$$\begin{bmatrix} X'X & X'Z_0 & X'Z_1 \\ Z_0'X & Z_0'Z_0 + A^{-1}\alpha_{00} & Z_0'Z_1 + A^{-1}\alpha_{01} \\ Z_1'X & Z_1'Z_0 + A^{-1}\alpha_{01} & Z_1'Z_1 + A^{-1}\alpha_{11} \end{bmatrix} \begin{pmatrix} \beta \\ a \\ \lambda \end{pmatrix} = \begin{pmatrix} X'y \\ Z_0'y \\ Z_1'y \end{pmatrix},$$

onde:

- $\alpha_{ij}$   $g^{ij} \sigma_e^2$
- $g^{ij}$  provém de  $G^{-1}$
- $\sigma_e^2$  variância residual média
- $\beta$  vetor de efeitos fixos

Assim:

$$G^{-1} \sigma_e^2 = \begin{pmatrix} \alpha_{00} & \alpha_{01} \\ \alpha_{10} & \alpha_{11} \end{pmatrix}$$

### 2.3 Estimação dos Componentes de Variância e Covariância

As variâncias genética e fenotípica são dependentes da idade, ou seja, podem aumentar ou diminuir com a idade. Tem-se:

$\hat{\sigma}_{A(i)}^2 = K_i' \hat{G} K_i = g_{aa} + 2A^* g_{a\lambda} + A^{*2} g_{\lambda\lambda}$  é a variância genética na idade  $i$ .

$\hat{\sigma}_{F(i)}^2 = \hat{\sigma}_{A(i)}^2 + \hat{\sigma}_e^2$  é a variância fenotípica na idade  $i$ .

$\hat{\sigma}_{A(ij)} = K_i' \hat{G} K_j = g_{aa} + (A_i^* + A_j^*) g_{a\lambda} + A_i^* A_j^* g_{\lambda\lambda}$  é a covariância genética entre as idades  $i$  e  $j$ .

$r_{A(ij)} = \frac{\hat{\sigma}_{A(ij)}}{\hat{\sigma}_{A(i)} \hat{\sigma}_{A(j)}}$  é a correlação genética entre as

idades  $i$  e  $j$ .

$K_i' = [1 \text{ idade } i]$  é o transposto do vetor que contempla o valor 1 e o valor da idade  $i$ .

Para a utilização das equações de modelo misto e estimação das variâncias, covariâncias e correlações genéticas são necessárias as estimativas de  $G$  e  $\sigma_e^2$ .

Meyer (1997), Shah et al. (1997) e Meyer & Hill (1997) reportam aspectos da estimação de parâmetros genéticos para dados longitudinais por máxima verossimilhança restrita (REML). Com base no algoritmo EM tem-se os seguintes estimadores para os componentes de variância:

$$\hat{\sigma}_e^2 = \frac{y'y - \hat{\beta}'X'y - \hat{a}'Z_0'y - \hat{\lambda}'Z_1'y}{N - r(x)}$$

$$\hat{g}_{ij} = \frac{(\hat{\theta}_i A^{-1} \hat{\theta}_j + tr A^{-1} C_{ij} \sigma_e^2)}{q}$$

onde:

$\theta_i$ :  $a_i$  ou  $\lambda_i$ .

$N$ : número total de observações.

$r(x)$ : posto de  $X$  ou número de colunas linearmente independentes de  $X$ .

- $q$ : número de elementos aleatórios (número de indivíduos ou valores genéticos a serem preditos).
- $tr$ : operação traço matricial ou soma dos elementos da diagonal da matriz.
- $C_{ij}$ : submatriz que provém da inversa generalizada ( $C$ ) da matriz dos coeficientes das equações de modelo misto.

$$C = \begin{bmatrix} C_{xx} & C_{xz_0} & C_{xz_1} \\ C_{z_0x} & C_{z_0z_0} & C_{z_0z_1} \\ C_{z_1x} & C_{z_1z_0} & C_{z_1z_1} \end{bmatrix}$$

Partindo-se de valores iniciais para  $G$  e  $\sigma_e^2$  obtém-se  $\beta$ ,  $a$  e  $\lambda$  a partir das equações de modelo misto, os quais são utilizados nos estimadores  $\hat{\sigma}_e^2$  e  $\hat{g}_{ij}$ , cujas estimativas são retornadas às equações de modelo misto. Procede-se desta maneira sucessivamente até a convergência.

Valores iniciais de  $g_{aa}$ ,  $g_{a\lambda}$  e  $g_{\lambda\lambda}$  podem ser obtidos a partir de estimativas de  $\sigma_{A(i)}^2$  para três idades, obtidas pelo método de máxima verossimilhança restrita.

#### 2.4 Predição dos Valores Genéticos em cada Idade

Os valores genéticos para o indivíduo  $j$  podem ser preditos para várias idades ( $i$ ) através da regressão linear aleatória:

$$\hat{V}G_{ij} = \hat{a}_j + \hat{\lambda}_j \text{ (idade } i\text{)}$$

As predições  $\hat{V}G_{ij}$  permitem a re-ordenação dos candidatos à seleção, nas diferentes idades, possibilitando a re-seleção de indivíduos de acordo com a idade desejada.

#### 2.5 Exemplo Aplicado a Dados Simulados

O presente exemplo considera a avaliação de 7 indivíduos (alguns avaliados em mais de uma idade) pertencentes a 2 populações, para o

caráter densidade da madeira em diferentes idades, conforme especificado na Tabela 1.

Tabela 1 – Genealogia e dados referentes a avaliação de 7 indivíduos, para o caráter densidade da madeira, em diferentes idades.

Indivíduo	Pai	Mãe	População	Idade (meses)	Densidade da madeira (kg/0,1m <sup>3</sup> )
4	1	-	1	4	20
4	1	-	1	15	35
5	3	2	2	8	27
5	3	2	2	10	30
6	1	2	2	6	21
6	1	2	2	12	32
7	4	5	2	13	28
7	4	5	2	15	32

A seguinte estrutura de variâncias e covariâncias foi considerada:

$$\sigma_e^2 = 5,50$$

$$G = \begin{bmatrix} 2,7500 & -0,0581 \\ -0,0581 & 0,0550 \end{bmatrix}$$

$$G^{-1}\sigma_e^2 = \begin{bmatrix} 2,0457 & 2,1610 \\ 2,1610 & 102,2828 \end{bmatrix}$$

## 2.6 Aplicação a dados experimentais

Foram utilizados dados do caráter circunferência à altura do peito, obtidos a partir de um teste de 45 progênies de *Eucalyptus urophylla*, conduzido no município de Aracruz-ES, pela empresa Aracruz Celulose S.A. O delineamento experimental empregado foi o de blocos ao acaso com 9 repetições e 6 plantas por parcela e o referido caráter foi avaliado em 7 idades (1 a 7 anos). Foram utilizados 14.278 dados, os quais foram analisados pelo programa DXMRR (Meyer, 1998b), disponível no aplicativo DFREML.

Adotou-se um modelo incluindo os efeitos aditivos (1 função de covariância com 3 componentes de variância:  $g_{aa}$ ,  $g_{a\lambda}$  e  $g_{\lambda\lambda}$ ), os efeitos



de ambiente permanente (1 função de covariância com 3 componentes de variância) e os efeitos residuais (7 componentes de variância do resíduo, ou seja, um para cada idade). Assim, o modelo de regressão aleatória esteve associado a 13 componentes de variância.

### 3 Resultados e discussão

#### 3.1 Exemplo numérico com dados simulados

Com base nos componentes de variância apresentados tem-se na Tabela 2, as estimativas de variância genética aditiva ( $\sigma_A^2$ ), variância fenotípica ( $\sigma_F^2$ ) e herdabilidade ( $h^2$ ) em função da idade.

Tabela 2 – Estimativas da variância genética aditiva ( $\sigma_A^2$ ), variância fenotípica ( $\sigma_F^2$ ) e herdabilidade ( $h^2$ ) em função da idade.

Idade	$\sigma_A^2$	$\sigma_F^2$	$h^2$
4	3,1652	8,6652	0,3653
5	3,5440	9,0440	0,3919
6	4,0328	9,5328	0,4230
7	4,6316	10,1316	0,4571
8	5,3404	10,8404	0,4926
9	6,1592	11,6592	0,5283
10	7,0880	12,5880	0,5631
11	8,1268	13,6268	0,5964
12	9,2756	14,7756	0,6278
13	10,5344	16,0344	0,6570
14	11,9032	17,4032	0,6840
15	13,3820	18,8820	0,7087

Verifica-se no presente exemplo, que a variação genética e a herdabilidade aumentam com o passar da idade, refletindo a mudança contínua do caráter em função do tempo, para a faixa de idades considerada.

Na Tabela 3 são apresentadas as covariâncias ( $\sigma_{A_{ij}}$ ) e correlações genéticas ( $r_{A_{ij}}$ ) entre alguns pares de idades.

Tabela 3 – Estimativas das covariâncias ( $\sigma_{A_{ij}}$ ) e correlações genéticas ( $r_{A_{ij}}$ ) entre alguns pares de idades.

Par de Idades	Covariância Genética ( $\sigma_{A_{ij}}$ )	Correlação Genética ( $r_{A_{ij}}$ )
4-5	3,3271	0,9934
4-15	4,9461	0,7600
14-15	12,6151	0,9995

Verificam-se altas correlações genéticas entre todas as combinações de idades, sendo mais altas as estimativas envolvendo medidas adjacentes, conforme esperado.

As matrizes de incidência são:

$$X = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 0 & 15 \\ 0 & 1 & 8 \\ 0 & 1 & 10 \\ 0 & 1 & 6 \\ 0 & 1 & 12 \\ 0 & 1 & 13 \\ 0 & 1 & 15 \end{bmatrix}, \text{ considerando um modelo de médias.}$$

$$Z_0 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Z_I = \begin{bmatrix} 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 15 \end{bmatrix}$$

A matriz  $A^{-1}$  equívale a:

$$A^{-1} = \begin{bmatrix} 1,8333 & 0,5000 & 0 & -0,6667 & 0 & -1,0000 & 0 \\ 0,5000 & 2,0000 & 0,5000 & 0 & -1,0000 & -1,0000 & 0 \\ 0 & 0,5000 & 1,5000 & 0 & -1,0000 & 0 & 0 \\ -0,6667 & 0 & 0 & 1,8333 & 0,5000 & 0 & -1,0000 \\ 0 & -1,0000 & -1,0000 & 0,5000 & 2,5000 & 0 & -1,0000 \\ -1,0000 & -1,0000 & 0 & 0 & 0 & 2,0000 & 0 \\ 0 & 0 & 0 & -1,0000 & -1,0000 & 0 & 2,0000 \end{bmatrix}$$

Resolvendo as equações de modelo misto obtém-se as soluções, conforme a Tabela 4.

Tabela 4 – Soluções das equações de modelo misto.

Efeitos	Soluções
Fixos	
$\hat{P}_1 + \hat{b}_0$	15,3375
$\hat{P}_2 + \hat{b}_0$	13,5377
$\hat{b}_1$	1,4179
Aleatórios a	
$\hat{a}_1$	-0,2663
$\hat{a}_2$	0,2006
$\hat{a}_3$	0,2663
$\hat{a}_4$	-0,4341
$\hat{a}_5$	0,4998
$\hat{a}_6$	-0,0986
$\hat{a}_7$	-0,3443

Continuação

Efeitos	Soluções
Aleatórios $\lambda$	
$\hat{\lambda}_1$	-0,0252
$\hat{\lambda}_2$	0,0529
$\hat{\lambda}_3$	0,0252
$\hat{\lambda}_4$	-0,0919
$\hat{\lambda}_5$	0,0642
$\hat{\lambda}_6$	0,0416
$\hat{\lambda}_7$	-0,1425

Com base nestas soluções tem-se as seguintes estimativas:

Média da população 1 aos 4 meses:

$$P_1 + b_0 + b_1 A^* = 15,3375 + 1,4179 \times 4 = 21,0091$$

Média da população 1 aos 15 meses:

$$P_1 + b_0 + b_1 A^* = 15,3375 + 1,4179 \times 15 = 36,6060$$

Média da população 2 aos 4 meses:

$$P_2 + b_0 + b_1 A^* = 13,5377 + 1,4179 \times 4 = 19,2093$$

Média da população 2 aos 15 meses:

$$P_2 + b_0 + b_1 A^* = 13,5377 + 1,4179 \times 15 = 34,8062$$

Os valores genéticos dos indivíduos, em diferentes idades, são apresentados na Tabela 5 e são dados por:

$\underline{VG}_1$  à idade de 4 meses:

$VG_{.14} = \hat{a}_1 + \hat{\lambda}_1 \times 4 = -0,2663 + (-0,0252) \times 4 = -0,3671$  e assim sucessivamente.

A metodologia permite então a obtenção de curvas de valores genéticos para cada indivíduo em função da idade, bem como a comparação dos vários indivíduos a uma idade comum, mesmo quando alguns indivíduos são avaliados em uma ou poucas idades, no intervalo de idades considerado.

Tabela 5 – Valores genéticos preditos dos indivíduos em diferentes idades.

Indivíduo	Valor Genético			
	Idade = 4 meses	Idade = 8 meses	Idade = 12 meses	Idade = 15 meses
1	-0,3671	-0,4679	-0,5687	-0,6443
2	0,4122	0,6238	0,8353	0,9940
3	0,3670	0,4676	0,5683	0,6437
4	-0,8018	-1,1695	-1,5372	-1,8130
5	0,7565	1,0133	1,2701	1,4626
6	0,0678	0,2342	0,4006	0,5254
7	-0,9141	-1,4840	-2,0538	-2,4811

### 3.2 Resultados experimentais

Foram comparadas 3 classes de modelos de análise: modelos univariados para cada idade separadamente; modelos de repetibilidade envolvendo as várias idades; modelos de regressão aleatória. Os resultados referentes aos modelos univariados são apresentados na Tabela 6.

Tabela 6 – Estimativas\* de parâmetros genéticos em análises univariadas para o caráter circunferência à altura do peito em *Eucalyptus urophylla* em várias idades, bem como o número ( $N$ ) de indivíduos avaliados em cada idade, a sobrevivência ( $S$ ) e as repetibilidades ( $r_{ij}$ ) para o caráter sobrevivência envolvendo as idades  $i$  e  $j$ .

Idade	Herdabilidade	Variância aditiva	N	S	i-j	$r_{ij}$
1	0,150 ± 0,04	2,6665	2326	0,9572	1-2	0,97
2	0,186 ± 0,05	9,8157	2255	0,9280	2-3	0,95
3	0,237 ± 0,06	18,3412	2144	0,8823	3-4	0,94
4	0,357 ± 0,09	36,7589	2017	0,8300	4-5	0,96
5	0,369 ± 0,09	48,8212	1936	0,7967	5-6	0,95
6	0,378 ± 0,09	54,9871	1839	0,7568	6-7	0,96

Continuação

Idade	Herdabilidade	Variância aditiva	N	S	i-j	r <sub>ij</sub>
7	0,390 ± 0,09	78,1604	1767	0,7272	1-7	0,76
					2-7	0,78
					3-7	0,82
					4-7	0,88
					5-7	0,91
					6-7	0,96

\* Obtidas pelo método da máxima verossimilhança restrita (REML).

Verifica-se pela Tabela 6, um aumento progressivo da herdabilidade e uma redução progressiva da sobrevivência com o aumento da idade.

Na Tabela 7 são apresentados os resultados associados ao modelo de regressão aleatória.

Verifica-se pela Tabela 7 que a herdabilidade aumenta ligeiramente com o passar da idade, praticamente se estabilizando entre os 3 e 5 anos de idade. As estimativas obtidas para as idades de 1, 2 e 3 anos são próximas às obtidas pelas análises univariadas (Tabela 6) mas as obtidas para as demais idades não o são.

A provável explicação para estas diferenças refere-se ao efeito da reduzida sobrevivência a partir (a menor repetibilidade para a sobrevivência em idades adjacentes, ocorreu para a combinação 3-4 anos, conforme a Tabela 6) do 4º ano e seus reflexos nas metodologias de análise. Por exemplo, para o 4º ano, a técnica da regressão aleatória baseou-se em 14.278 dados de 2.326 indivíduos, ao passo que a análise univariada baseou-se em apenas 2.017 dados de 2.017 indivíduos. Assim, a análise univariada estima parâmetros para uma população com reduzida variância fenotípica devido à morte dos indivíduos menos vigorosos. Por outro lado, a técnica da regressão aleatória utiliza a correlação entre idades para aproveitar a informação dos indivíduos que morreram em idades prévias, estimando assim, os parâmetros para a população inicial como um todo e não para a população já selecionada naturalmente para adaptação. Estes efeitos da sobrevivência nas estimativas de parâmetros genéticos foram também relatados por Matheson & Raymond (1984) em *Pinus radiata* e também através de simulação. Tais autores relataram que as estimativas de herdabilidade para as populações completas foram de 0,12 e 0,24 e

passaram a 0,21 e 0,33 após a eliminação das piores plantas, para 2 populações de *Pinus radiata*, respectivamente.

Tabela 7 – Estimativas\* de parâmetros genéticos pelo modelo de regressão aleatória para o caráter circunferência à altura do peito em *Eucalyptus urophylla* nas idades de 1 a 7 anos.

Idade	Herdabilidade	Variância aditiva	Par de idades	Correlação genética	Repetibilidade
1	0,153 ± 0,02	2,7475	1-2	0,862	0,909
2	0,176 ± 0,03	8,9071	1-3	0,756	0,868
3	0,211 ± 0,03	21,3117	1-4	0,697	0,834
4	0,232 ± 0,04	39,9610	1-5	0,661	0,811
5	0,245 ± 0,05	64,8552	1-6	0,637	0,790
6	0,251 ± 0,05	95,9942	1-7	0,620	0,767
7	0,251 ± 0,05	133,3780	2-3	0,983	0,947
			2-4	0,964	0,940
			2-5	0,950	0,933
			2-6	0,939	0,921
			2-7	0,932	0,904
			3-4	0,996	0,981
			3-5	0,991	0,981
			3-6	0,986	0,973
			3-7	0,982	0,959
			4-5	0,999	0,994
			4-6	0,997	0,989
			4-7	0,995	0,976
			5-6	0,999	0,996
			5-7	0,999	0,984
			6-7	1,0	0,982

\* Baseada em 14.278 dados de 2.326 indivíduos.

Dessa forma, os resultados obtidos por modelos univariados para cada idade não são precisos. À medida em que os anos passam, os parâmetros passam a ser representativos de uma população selecionada para adaptação e são, portanto, superestimados conduzindo a infe-

rências incorretas sobre os valores genéticos e ganhos genéticos. Entretanto, se avaliações de idades precoces (2 ou 3 anos) são utilizadas para a seleção, os erros pelo uso dos modelos univariados provavelmente são desprezíveis (Tabelas 6 e 7).

Um caráter avaliado repetidas vezes sobre um mesmo indivíduo pode ser considerado como um mesmo caráter ao longo das idades se as correlações genéticas entre medidas repetidas deste caráter for igual a 1 (Henderson, 1984). As estimativas de correlações genéticas apresentadas na Tabela 7 indicam que a circunferência à altura do peito pode ser considerada um mesmo caráter a partir dos três anos (correlações > 0,982), idade esta que coincide com o início da estabilização da herdabilidade do caráter. Pode-se inferir, com base nas correlações, que o crescimento ao longo das idades equivale praticamente a 3 caracteres: crescimento a 1 ano; crescimento aos 2 anos; crescimento a partir de 3 anos. Os resultados referentes às estimativas de repetibilidade (correlações fenotípicas entre idades) apresentados na Tabela 7 e incrementos médios anuais (Tabela 8) corroboram esta inferência.

As estimativas de repetibilidade (Tabela 7) indicam que um modelo de repetibilidade pode ser adequado para dados de crescimento a partir dos 3 anos (pois a amplitude das estimativas de repetibilidade é reduzida para esta faixa de idades). Entretanto, para adoção do modelo de repetibilidade é mais adequado considerar a variável incremento anual. Resultados referentes aos modelos de repetibilidade para o incremento anual da circunferência são apresentados na Tabela 8.

Verifica-se que tratando o incremento anual na circunferência como um único caráter, o mesmo apresentou herdabilidade de 0,18 e repetibilidade de 0,67. Entretanto tal variável não deve ser tratada como um único caráter já que herdabilidades e repetibilidades diferentes foram obtidas nas faixas de idade de 1 a 2 anos e de 3 a 7 anos. Os valores de incremento médio anual também são mais próximos a partir do 3º incremento, sendo que nos dois primeiros anos a taxa de crescimento é muito superior (Tabela 8). Com base nestes resultados e naqueles da Tabela 7, existe evidência de que o modelo de repetibilidade pode ser aplicado adequadamente a partir do 3º incremento.

Em função da pequena importância da avaliação de 1 ano e do fato de que dados nesta idade raramente são tomados na prática, pode-se recomendar a predição dos valores genéticos em duas etapas: aos 2 anos para circunferência empregando-se a análise univariada; para incrementos anuais a partir do 2º ano pelo modelo de repetibilidade. Dessa forma, os valores genéticos preditos para a idade de 7 anos são



dados  $\hat{a}_2 + 5\hat{a}_m$ , em que  $\hat{a}_2$  referem-se aos valores genéticos preditos para circunferência aos 2 anos e  $\hat{a}_m$  referem-se aos valores genéticos preditos para o incremento anual.

Tabela 8 – Estimativas de parâmetros genéticos empregando-se modelos de repetibilidade para o caráter incremento anual da circunferência à altura do peito em *Eucalyptus urophylla*, bem como as médias de idade e incrementos médios anuais.

Idades envolvidas	Herdabilidade	Repetibilidade	Variância aditiva	Média (cm)	Incremento médio anual (cm)
1 a 7 anos	0,180 ± 0,04	0,67	2,7466	-	-
3 a 7 anos	0,210 ± 0,05	0,69	1,9249	-	-
1 e 2 anos	0,126 ± 0,04	0,57	2,2071	-	-
1	-	-	-	14,21	14,21
2	-	-	-	25,97	11,76
3	-	-	-	31,08	5,11
4	-	-	-	35,73	4,65
5	-	-	-	38,94	3,21
6	-	-	-	41,37	2,43
7	-	-	-	45,99	4,62

Em resumo, na análise quantitativa de medidas repetidas 4 classes de modelos de análise podem ser empregadas: (i) modelos univariados para cada idade; (ii) modelos de repetibilidade; (iii) modelos de regressão aleatória e (iv) modelos multivariados.

Destes, os mais completos são os modelos multivariados, que utilizam toda a informação simultaneamente e tratam cada idade como um caráter, considerando suas diferentes herdabilidades e correlações genéticas. Entretanto, tais modelos são superparametrizados e praticamente proibitivos na prática quando muitas idades são consideradas, tal como no presente trabalho. Isto ocorre devido ao fato de que, neste caso, problemas de estimação surgem principalmente quando o verdadeiro valor do parâmetro encontra-se próximo (como no presente trabalho em que as correlações foram muito próximas a 1) ao limite do espaço paramétrico (Meyer & Hill, 1997), acarretando problemas computacionais e produzindo estimativas imprecisas. Assim, devem ser preferidos modelos com menor número de parâmetros e que descrevam adequadamente os dados observados.

Os modelos de regressão aleatória podem ser considerados como modelos multivariados simplificados ou reduzidos, que permitem a obtenção dos mesmos parâmetros de interesse que podem ser obtidos pelos modelos multivariados, mas com uma menor parametrização. No presente trabalho, um modelo multivariado requereria 56 parâmetros (28 covariâncias aditivas, 21 covariâncias de ambiente permanente e 7 variâncias residuais), ao passo que a regressão aleatória empregou 13 parâmetros e, portanto, mostrou-se uma técnica muito informativa e adequada.

Com relação aos outros dois modelos, no presente trabalho o modelo de repetibilidade mostrou-se adequado para os dados a partir do 3º incremento anual e o modelo univariado mostrou-se adequado para dados até o 3º ano. Empregados nestas situações estes modelos, embora simplificados, são úteis e eficientes.

Os modelos mais completos (multivariado e de regressão aleatória) devem ser utilizados de acordo com o número de idades disponíveis para análise: com dados de 2 idades (por exemplo 2 e 3 anos), o modelo multivariado pode ser utilizado de maneira eficiente; com dados de 4 ou mais idades o modelo de regressão aleatória pode ser utilizado adequadamente.

#### **4 Conclusões**

- A técnica de regressão aleatória permite: descrever continuamente a estrutura de covariâncias de caracteres de crescimento ao longo do tempo; a estimação simultânea de parâmetros genéticos ao longo das idades, considerando a mudança genética do caráter ao longo do tempo; a predição de valores genéticos de indivíduos avaliados em diferentes idades e a projeção destes valores genéticos para uma idade comum para efeitos de ordenamento e seleção; a obtenção de curvas de valores genéticos para cada indivíduo em função da idade.
- O modelo de regressão aleatória na análise genética conduziu a melhores resultados do que os tradicionais modelos univariados e modelos de repetibilidade, devendo assim, ser preferido.
- O modelo de regressão aleatória permitiu a obtenção de estimativas dos mesmos parâmetros que podem ser obtidos por um modelo multivariado, porém com uma menor parametrização, fato que é vantajoso computacionalmente.

RESENDE, M. D. V. de; REZENDE, G. D. S. P.; FERNANDES, J.S.C.; Random regression and covariance functions in the analysis of repeated measures. *Rev. Mat. Est.* (São Paulo), v.19, p.21-40, 2001.

- **ABSTRACT:** *The random regression theory emphasizes that, in a experiment, different random regression coefficients are associated to different individuals of the population if those individuals are treated as random effects. The objectives of this paper were to present and study the random regression technique for the purpose of applications in the genetic improvement of perennial species. Using simulated and experimental data, this technique provided: the description of the covariance's structure of growth traits through the ages; the simultaneous estimation of genetic parameters through time, considering the genetic change of the trait with age; the breeding values prediction for individuals evaluated in different ages and the projection of these values to a common age for purposes of ranking and selection; the obtention of curves of genetic values for each individual as a function of age. The random regression model showed better results than the univariate and repeatability models and provided estimates of the same parameters that can be obtained by a multivariate method, however, with a lesser parameterization, which is computationally advantageous.*
- **KEYWORDS:** *Mixed models; multivariate models; linear prediction; repeated measures; longitudinal data; genetic selection.*

### Referências bibliográficas

HENDERSON JUNIOR, C. R. Analysis of covariance in the mixed model: higher-level, nonhomogeneous and random regression. *Biometrics*, v.38, p.633-40, 1982.

HENDERSON, C. R. *Applications of linear models in animal breeding*. Guelph: University of Guelph, 1984. 462p.

JAMROZIK, J., SCHAEFFER, L. R. Estimates of genetic parameters for a test day model with random regressions for yield of first lactation Holsteins. *J. Dairy Sci.*, v.80, p.726-70, 1997.

JAMROZIK, J., SCHAEFFER, L. R., DEKKERS, J. C. M. Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.*, v.80, p.1217-26, 1997.

KIRKPATRICK, M., LOFSVOLD, D., BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, v.124, p.979-93, 1990.

- KIRKPATRICK, M., HILL, W. G., THOMPSON, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactations in dairy cattle. *Genet. Res.*, v.64, p.57-69, 1994.
- LONGFORD, N. T. *Random coefficients models*. Oxford: Clarendon Press, 1993. 390p. (Oxford Statistical Science Series).
- MATHESON, A. C., RAYMOND, C. A. Effects of thinning in progeny tests on estimates of genetic parameters in *Pinus radiata*. *Silvae Genet.*, v.4-5, p.125-8, 1984.
- MEYER, L. An average information restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices. *Genet. Sel., Evol.*, v.29, p.97-116, 1997.
- MEYER, K. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.*, v.30, p.221-49, 1998a.
- MEYER, K. "DXMRR": A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 6., 1998b, Armidale. *Proceedings ...* v.27, p.465-6.
- MEYER, K., HILL, W. G. Estimation of genetic and phenotypic covariance functions for longitudinal or repeated records by restricted maximum likelihood. *Livest. Prod. Sci.*, v.47, p.185-200, 1997.
- SCHAEFFER, L. R., DEKKERS, J. C. M. Random regressions in animal models for test-day production in dairy cattle. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5., 1994, Edinburgh. *Proceedings ...* v.18, p.443.
- SHAH, A., LAIRD, N., SCHOENFELD, D. A random effects model for multiple characteristics with possible missing data. *J. Am. Stat. Assoc.*, v.92, p.775-9, 1997.

Recebido em 3.4.2000