



# LGN 5810 - Quantitative Genetics

---

## About the course

Antonio Augusto Franco Garcia (<http://statgen.esalq.usp.br>)

Department of Genetics, ESALQ/USP, 2018

# Modelos Mistos em Genética Quantitativa

*Camila Godoy, Cristiane Taniguti, Fernando Correr*

*November 25, 2018*

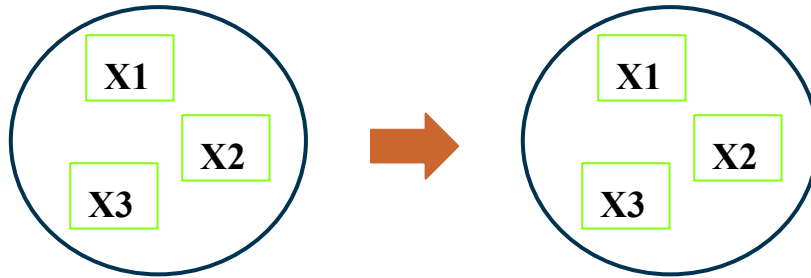
**Estimação dos valores genéticos (breeding values) de plantas não avaliadas utilizando modelos mistos e matriz de parentesco (pedigree).**

## **Conteúdo:**

- Revisão: Modelos Mistos e Matriz de Parentesco
- Cinco cenários diferentes para explicar como estimar os valores genéticos.

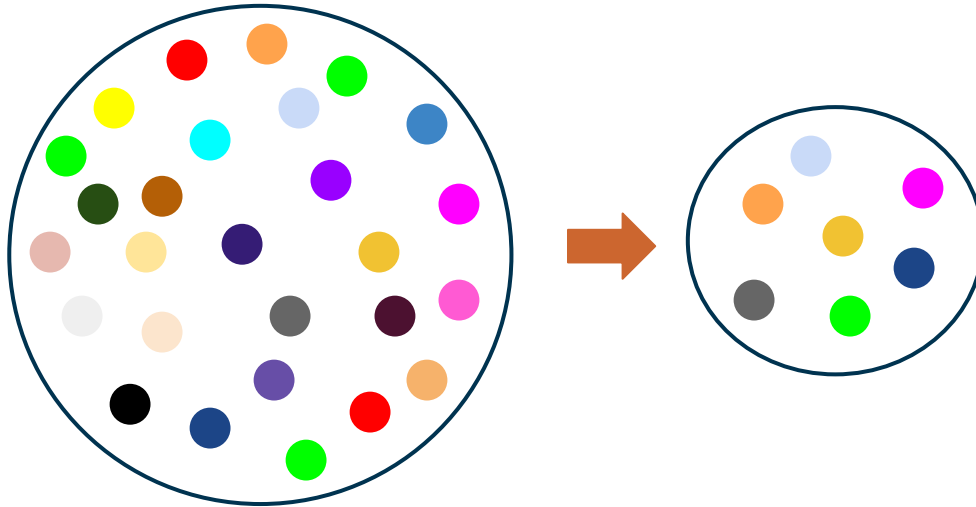
# MODELOS MISTOS

1. Modelos com efeitos fixos e aleatórios são denominados modelos mistos
2. Efeito fixo: efeitos particulares de cada nível de um determinado fator



# MODELOS MISTOS

3. Efeito aleatório: os níveis dos fatores representam uma amostra aleatória de uma população de todos os níveis de fatores possíveis



# EFEITOS FIXOS E EFEITOS ALEATÓRIOS

1. Se o parâmetro é uma constante fixa, e desejamos estimar, é um efeito fixo
2. Se o parâmetro é extraído de alguma distribuição de probabilidade, e alguém quer fazer inferências sobre a distribuição, ou realizações da distribuição, é um efeito aleatório

# BLUE E BLUP

1. Estimadores dos efeitos fixos: BLUE (Best Linear Unbiased Estimator)
2. Preditores dos efeitos aleatórios: BLUP (Best Linear Unbiased Predictor)



$k$  efeitos

Fixo:  $k$  graus de liberdade

Aleatório: um grau de liberdade para estimar a variância

**Modelos mistos podem ser usados em situações em que  $n < p$**

# MODELOS MISTOS

1. Amplamente utilizados em situações em que as respostas são agrupadas em torno de efeitos aleatórios
2. Dependência natural entre as observações no mesmo cluster (mesmo grupo) (por exemplo, indivíduos aparentados)
3. MM é uma forma natural de você considerar efeitos que são importantes na genética (grau de relacionamento entre os indivíduos)
4. Experimentos desbalanceados



# MODELO

$$y = X\beta + Zu + e$$

- $y$ : vetor de resposta (observações)
- $\beta$ : vetor de efeitos fixos
- $u$ : vetor de efeitos aleatórios
- $X$  e  $Z$ : matrizes de incidência conhecidas
- $e$ : vetor residual



# SUPOSIÇÕES

1.  $u \sim \text{MVN}(0, G)$

2.  $e \sim \text{MVN}(0, R)$

3.  $u$  e  $e$  são independentes:

$$\begin{pmatrix} u \\ e \end{pmatrix} \sim \text{MVN} \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \right\}$$

4.  $y \sim \text{MVN}(X\beta, V); V = ZGZ' + R$

# MATRIZ G E R

1. Diferentes estruturas de matriz de variância e covariância:

- identidade
- diagonal
- simetria composta
- não estruturada
- autoregressiva

**ASReml-R, o Genstat ou o SAS**

**Iremos focar na matriz G (útil em genética quantitativa/covariância)** 😊

# ESTIMAÇÃO DOS PARÂMETROS

## Função de verossimilhança

- The density function of  $y$  is

$$p(y|\theta, V) = (2\pi)^{-n/2}|V|^{-1/2}\exp\left\{-\frac{1}{2}(y - X\theta)'V^{-1}(y - X\theta)\right\}$$

- Likelihood

$$L(\theta, V) \propto |V|^{-1/2}\exp\left\{-\frac{1}{2}(y - X\theta)'V^{-1}(y - X\theta)\right\}$$

- Log-likelihood

$$l(\theta, V) = \log[L(\theta, V)] \propto -\frac{1}{2}\log|V| - \frac{1}{2}(y - X\theta)'V^{-1}(y - X\theta)$$



# ESTIMAÇÃO DOS PARÂMETROS

Usando essa abordagem, o melhor estimador linear não-viesado (BLUE) dos efeitos fixos é dado por:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

O melhor preditor linear não-viesado (BLUP) dos efeitos aleatórios é dado por:

$$\hat{u} = GZ' (ZGZ' + R)^{-1} (y - X\beta)$$



# EQUAÇÕES DE MODELOS MISTOS (MME) (Henderson (1950))

1. Estimação dos parâmetros que estão em  $\beta$  e a predição dos parâmetros que estão em  $u$  usando MME.
2. Maximizando a densidade conjunta de  $y$  e  $u$  para  $\beta$  e  $v$  (MME):

$$\begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix}$$

# EQUAÇÕES DE MODELOS MISTOS (MME) (Henderson (1950))

Usando o MME, pode ser mostrado que  $u$  e  $\beta$  podem ser obtidos por:

$$\hat{u} = (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1} (y - X \hat{\beta})$$

$$\hat{\beta} = \{X' [R^{-1} - R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1}] X\}^{-1}$$

$$X' [R^{-1} - R^{-1} Z (Z' R^{-1} Z + G^{-1})^{-1} Z' R^{-1}] y$$

# BLUP Breeding Values

1. Os programas de melhoramento baseiam-se no princípio de que indivíduos aparentados têm algum grau de associação.
2. A semelhança entre pais e filhos baseia-se no componente aditivo da variação genética. Portanto, este componente é de grande importância em programas de melhoramento.
3. **MM permite modelar na matriz  $G$  o parentesco ou relações entre os indivíduos que estamos usando no melhoramento.**



# ANIMAL MODEL

$$y = X\beta + Zu + e$$

$y$ : vetor de observações fenotípicas


$\beta$ : vetor de efeitos fixos

$u \sim \text{MVN}(0, G)$ : vetor dos breeding values, relativo a todos os indivíduos com registro ou no arquivo pedigree.

$e \sim \text{MVN}(0, I_n\sigma^2_e)$ : efeitos residuais.



# MATRIZ A

1. A matriz G descreve a covariância entre efeitos aleatórios. Aqui, esses efeitos são os breeding values
2. A covariância genética aditiva entre dois indivíduos aparentados i e i' é dada por  $2\theta_{ii'}\sigma_A^2$    $G = A\sigma_A^2$
3. A é a matriz de relação genética aditiva, com elementos  $a_{ii'} = 2\theta_{ii'}$

# COEFICIENTE DE PARENTESCO ( $\theta$ )

## Coefficients of Coancestry and Fraternity

Common relationships, no inbreeding ( $\Delta_1$  to  $\Delta_6 = 0$ )

- $\Theta_{xy} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$
- $\Delta_{xy} = \Delta_1 + \Delta_7$

Relationship	$\Delta_7$	$\Delta_8$	$\Delta_9$	$\Theta_{xy}$	$\Delta_{xy}$
Parent-offspring	0	1	0	$\frac{1}{4}$	0
Half-sibs	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	0
Full-sibs	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

# MME E ANIMAL MODEL

As equações de modelos mistos podem ser adaptados ao “animal model”. Considerando que:

$$G = A\sigma_A^2 \quad \longrightarrow \quad G^{-1} = A^{-1}\sigma_A^{-2}$$

$$R = I\sigma^2 \quad \longrightarrow \quad R^{-1} = I\sigma^{-2}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\cdot \quad \lambda = \frac{\sigma_e^2}{\sigma_A^2} = \frac{1-h^2}{h^2}$$

# BLUE E BLUP

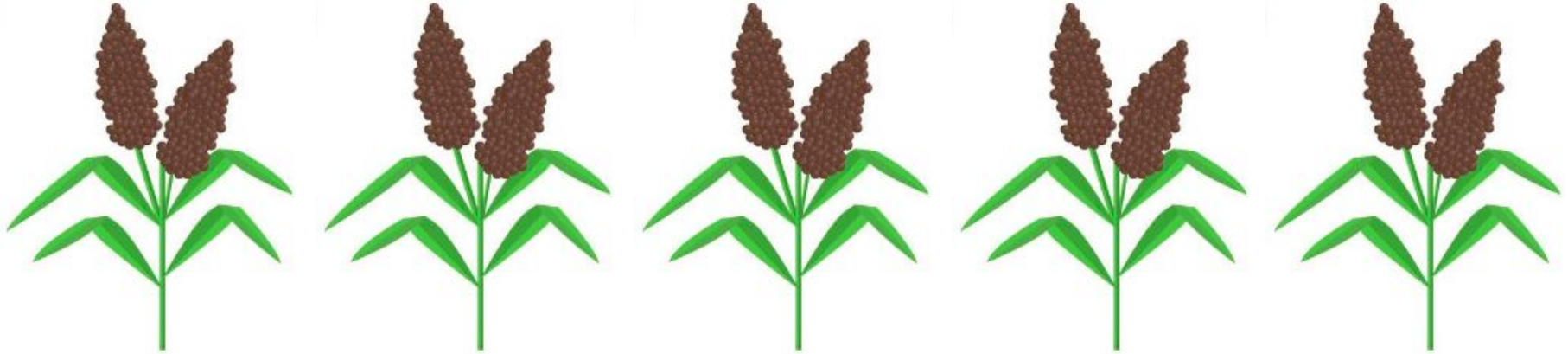
$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

$$\hat{u} = (Z'Z + \lambda A^{-1})^{-1} Z'(y - X\hat{\beta})$$

- Estimated breeding values (EBV)
- Indivíduos com maiores EBVs são escolhidos para a próxima geração
- Acurácia preditiva:  $(\rho(u, \hat{u})) \rightarrow$  correlação entre os valores do BLUPs com os valores observados.

# ESTIMAÇÃO DOS VALORES GENÉTICOS

## Exemplos práticos



Característica fenotípica:  
peso de 1000 sementes  
média: 26g

