



10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016  
02 a 04 de agosto de 2016 – Campinas, São Paulo  
ISBN 978-85-7029-135-6

## IMPLEMENTAÇÃO DE ALGORITMO PARA AVALIAÇÃO GENÉTICA DE GRANDES POPULAÇÕES DE ANIMAIS DE INTERESSE AGROPECUÁRIO

Daniel Henrique **Barboza**<sup>1</sup>; Caio Augusto **Volpato** Cunha<sup>2</sup>; Roberto Hiroshi **Higa**<sup>3</sup>

Nº 16602

### RESUMO

*Os programas de melhoramento de gado de corte atualmente coordenados pela Embrapa utilizam software proprietário, desenvolvido por universidades estrangeiras, para a estimação do valor genético dos animais. O projeto MaxiDep propõe o desenvolvimento de um ferramental próprio da Embrapa para avaliação genética de grandes populações animais para utilização por seus programas de melhoramento genético.*

*Este trabalho é parte do Plano de Ação 4, responsável por propôr e desenvolver uma solução computacional que contemple os modelos estatísticos atualmente utilizados para estimar valor genético de animais. Apresentamos uma versão inicial da solução em desenvolvimento para estimação dos efeitos desejados através do modelo animal univariado, utilizando duas abordagens distintas para a obtenção do melhor estimador linear não viesado (BLUP) dos parâmetros do modelo.*

**Palavras-chaves:** Sistemas de Equações Lineares; Modelo Animal; Modelos Lineares Mistos; Melhoramento Animal; Python.

1 Autor, Bolsista CNPq (PIBIC): Graduação em Estatística, IMECC / Unicamp, Campinas-SP; hbdaniel@outlook.com

2 Bolsista Embrapa: Graduação em Matemática Aplicada e Computacional, IMECC / Unicamp, Campinas-SP.

3 Orientador: Pesquisador da Embrapa Informática Agropecuária, Campinas-SP; roberto.higa@embrapa.br



**10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016**  
**02 a 04 de agosto de 2016 – Campinas, São Paulo**  
**ISBN 978-85-7029-135-6**

## **ABSTRACT**

*The programs of beef cattle improvement currently coordinated by Embrapa are using proprietary software, developed by foreign universities, for estimation of breeding value. The MaxiDep project propose the development of an own Embrapa's tooler for animal breeding in big populations of animals, for use by genetic improvement programs.*

*This work is part of Action Plan 4, which is responsible for propose and develop a computational solution contemplating the statistical models currently in use to estimate genetic effect of an animal. We present a initial version of solution under development to estimate effects through univariate animal model, using two different approaches to get the best linear unbiased predictor (BLUP).*

**Keywords:** System of Linear Equations; Animal Model; Linear Mixed Models; Animal Breeding; Python.

## **1 INTRODUÇÃO**

Para estimar a contribuição genética de animais para características fenotípicas de interesse, é comum utilizar-se modelos lineares mistos (Demidenko, 2004), em que são considerados diversos efeitos fixos e aleatórios, sendo a contribuição genética modelada como um efeito aleatório.

O método mais difundido para estimação de parâmetros neste tipo de modelo é o Melhor Preditor Linear não Viesado (BLUP) (Robinson, G.K., 1991). Um estimador é dito não viesado quando seu valor esperado coincide com a variável que está sendo estimada. O BLUP é a função linear não viesada das observações que minimiza a variância do estimador, portanto é melhor sob este critério. O processo para obtenção do BLUP envolve a resolução de um sistema de equações lineares, conhecidos como equações do modelo misto (Mixed Model Equations, MME).

Em programas de melhoramento genético animal, o MME resultante pode ter milhões de



**10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016**  
**02 a 04 de agosto de 2016 – Campinas, São Paulo**  
**ISBN 978-85-7029-135-6**

equações, variando em função do número de registro de animais em avaliação, o que faz com que sua solução por métodos como eliminação gaussiana ou decomposições tais como LU ou Cholesky não sejam viáveis. Em contrapartida, a matriz de coeficientes do MME é esparsa, constituindo-se numa característica a ser explorada na construção de soluções customizadas.

Em geral, para resolução das MME são utilizados métodos iterativos como Jacobi, Gauss-Seidel, sobre-relaxação sucessiva, gradientes conjugados e gradientes conjugados preconditionados (Miztal, 2011). Dada a magnitude do sistema linear resultante, muitas vezes esses algoritmos são implementados sem que o sistema de equações seja construído na memória, mas carregando na mesma apenas os dados necessários para a execução de uma iteração. Essa técnica é conhecida como Iteração nos Dados (IoD) (Schaeffer, 1986).

O método IoD preza pela possibilidade de trabalhar com um grande volume de dados, mantendo em memória apenas a solução atual, alternando entre a atualização das estimativas dos efeitos fixos e dos efeitos aleatórios, considerando toda a estrutura de correlação genética entre os animais. Em princípio, é necessário apenas um pré-processamento dos dados para deixá-los de forma adequada para atualização sucessiva dos valores resposta. Como ponto negativo, implementações de IoD são construídas com base em um modelo específico e, por essa razão, são menos flexíveis quanto à sua aplicabilidade a diferentes modelos.

Já os métodos referenciados como gerais, possuem a vantagem de resolver problemas para uma gama maior de modelos, por vezes em um tempo menor que o do IoD, porém há a limitação de recurso computacional, essencialmente memória RAM, já que é necessário manter um grande número de equações em memória.

Este trabalho apresenta resultados preliminares e discussões sobre a estimação de efeitos de interesse por meio do BLUP, através de duas abordagens para a resolução do sistema linear esparsa: IoD e Gradientes Conjugados.



## 2 MATERIAL E MÉTODOS

Toda a implementação dos métodos computacionais utilizados para a resolução do problema apresentado foi feita em linguagem Python.

As seguintes bibliotecas foram utilizadas:

- Numpy
- SciPy
- pyTables

O modelo misto adotado é descrito na equação abaixo:

$$y = Xb + Za + \varepsilon$$

onde o vetor de observações  $y$ ,  $X$ ,  $b$ ,  $Z$ ,  $a$  e  $\varepsilon$  são, respectivamente, vetor de observações, matriz de desenho dos efeitos fixos, vetor de efeitos fixos, matriz de efeitos aleatórios, vetor de efeitos animais e o vetor de resíduos.

Para um conjunto de  $n$  registros de variáveis respostas dos animais, considerando um total de  $q$  níveis dentre todos os fatores fixos, define-se:

$$y = [y_1, \dots, y_n]$$

$$X = [x_1, \dots, x_n]'$$

$$x_i = [x_{i1}, \dots, x_{iq}], i = 1, \dots, n$$

$$b = [b_1, \dots, b_q]$$

$$Z = [z_1, \dots, z_n]'$$

$$z_i = [z_{i1}, \dots, z_{iq}], i = 1, \dots, n$$

$$a = [a_1, \dots, a_n]$$

$$\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]$$



**10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016**  
**02 a 04 de agosto de 2016 – Campinas, São Paulo**  
**ISBN 978-85-7029-135-6**

Para estimação dos vetores  $b$  e  $a$ , utilizou-se o BLUP, cujas estimativas são obtidas pela resolução do sistema linear **(1)**, conhecido como Equações do Modelo Misto (MME):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (1)$$

onde  $\alpha$  é variância genética e  $A$  é a matriz de correlação genética, chamada de matriz de parentesco (Mrode, 1996).

Para a construção do MME, foram implementados algoritmos para a montar a matriz de parentesco e sua inversa, a partir do pedigree (banco de dados), bem como funções auxiliares para manipulação dos dados utilizados tanto no método de Gradientes Conjugados, quanto no de IoD.

Após extrair as informações do banco de dados original, há atualmente duas maneiras de obter as estimativas:

- (i) por solução direta do sistema linear, através do método dos gradientes conjugadores.
- (ii) através do algoritmo de interação nos dados (IoD).

O método (i) foi implementado com apoio da biblioteca SciPy, especialmente no que se refere ao armazenamento e organização da matriz esparsa para a montagem do sistema linear, além do próprio algoritmo do método de gradientes conjugados.

O método (ii) foi implementado sem auxílio de bibliotecas externas e utiliza o sistema de atualização dos valores baseado no algoritmo de Gauss-Seidel (Mrode, 1996). Algumas modificações em relação ao algoritmo tradicional foram introduzidas com o intuito de reduzir o custo computacional e o tempo, considerando a estrutura criada previamente no pré-processamento do pedigree.

Todas as informações extraídas do pedigree e também a estimativa do BLUP é salva em um arquivo HDF5, o qual possui estrutura similar à um banco de dados, com alto nível de compressão. Para tal, são utilizadas a biblioteca pyTables.



### 3 Resultados e Discussão

Como teste dos métodos, foi utilizado o exemplo 3.1, apresentado em (Mrode, 1996), onde a característica analisada é o ganho de peso à desmama, do inglês Pre-weaning weight gain (WWG). Os dados são reproduzidos na Tabela 1.

**Tabela 1.** Pedigree e ganho de peso à desmama de bezerros de raça.

Bezerro	Sexo	Pai	Mãe	WWG (kg)
4	Macho	1	Desconhecido	4.5
5	Fêmea	3	2	2.9
6	Fêmea	1	2	3.9
7	Macho	4	5	3.5
8	Macho	3	6	5.0

Adotou-se o seguinte modelo para estimar o efeito do fator sexo e da genética do bezerro:

$$y_{ij} = b_i + a_j + \varepsilon_{ij}$$

No qual  $y_{ij}$  é o WWG do  $j$ -ésimo bezerro do  $i$ -ésimo sexo;  $b_i$  é o do sexo;  $a_j$  é o efeito genético do animal;  $\varepsilon_{ij}$ , o erro aleatório associado a cada  $y_{ij}$ .

A obtenção do BLUP deu-se pela resolução do sistema linear **(1)**, com as matrizes correspondentes aos dados da Tabela 1, considerando  $\alpha = 2$ .

Vemos na Tabela 2 que as soluções encontradas pelos métodos (i) e (ii) condizem com as apresentadas em (Mrode, 1996).



**10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016**  
**02 a 04 de agosto de 2016 – Campinas, São Paulo**  
**ISBN 978-85-7029-135-6**

**Tabela 2.** Estimativas BLUP dos efeitos de sexo e do valor genético de bezerros de raça obtidas por Mrode, IoD e Gradientes Conjugados.

<b>Efeito</b>	<b>Solução (Mrode)</b>	<b>Estimativa - IoD</b>	<b>Estimativa – Gradientes Conjugados</b>
<b>Sexo</b>			
<b>Macho</b>	4.358	4.358	4.358
<b>Fêmea</b>	3.404	3.404	3.404
<b>Genético/Animal</b>			
<b>1</b>	-0.041	-0.041	-0.041
<b>2</b>	-0.091	-0.091	-0.091
<b>3</b>	0.098	0.098	0.098
<b>4</b>	-0.009	-0.009	-0.009
<b>5</b>	-0.186	-0.186	-0.186
<b>6</b>	0.177	0.177	0.177
<b>7</b>	-0.249	-0.249	-0.249
<b>8</b>	0.183	0.183	0.183

#### **4 Conclusão**

Viu-se que ambas implementações propostas resolveram corretamente o problema abordado na referência.

O projeto mostrou-se promissor por implementar com sucesso os métodos propostos em linguagem Python, mostrando a viabilidade de se expandir a solução a partir dos resultados obtido até agora.

Para o prosseguimento do projeto, planeja-se a expansão da quantidade de modelos, com generalizações que envolvem modelos multivariados e respostas de natureza categórica, implementação do IoD baseado no algoritmo de gradientes conjugados, além da incorporação de estruturas de covariâncias mais complexas.



**10º Congresso Interinstitucional de Iniciação Científica – CIIC 2016**  
**02 a 04 de agosto de 2016 – Campinas, São Paulo**  
**ISBN 978-85-7029-135-6**

### **Agradecimentos**

Ao meu querido avô, Orlando Barboza, por tudo o que fez por mim enquanto em vida.

À minha mãe, Amélia Cristina Barboza e à minha avó, Juliana Paes Barboza, pelo apoio.

Ao meu orientador, Roberto Hiroshi Higa, por ter me dado a oportunidade de participar do projeto e também por seu empenho, dedicação, paciência, compreensão e disponibilidade.

Ao meu amigo Caio Volpato, por toda a colaboração durante a elaboração do projeto, especialmente na parte computacional.

Aos professores do IMECC, por todo o conhecimento transmitido durante os últimos anos.

À Embrapa Informática Agropecuária e Embrapa Gado de Corte por fornecer toda a estrutura necessária.

Ao CNPq pelo suporte financeiro ao projeto.

### **Referências**

DEMIDENKO, E. *Mixed models: theory and applications.*; Hoboken, NJ: Wiley, 2004, 752p.

MISZTAL, I. *Computational techniques in animal breeding. University of Georgia. 2011. Notas de aula.*

MRODE, R.A., THOMPSON, R. *Linear Models for the Prediction of Animal Breeding Value*, Wallingford, CABI Publishing, 1996, 344p.

ROBINSON, G.K. That BLUP is a Good Thing: The Estimation of Random Effects. *Statist. Sci.* 6, n. 1, p. 15–32, 1991.

SCHAEFFER L, R., KENNEDY B. W., *Computing Strategies for Solving Mixed Model Equation. Journal of Dairy Science*, Guelph, Ontario, p. 575–579, fev. 1986.