



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

**Evelyne de Farias Siqueira**

**APLICAÇÃO DE MODELOS MISTOS E MODELOS LINEARES  
GENERALIZADOS MISTOS NA ESTIMAÇÃO DE COMPONENTES DE  
VARIÂNCIA**

Campina Grande  
Fevereiro de 2014

Evelyne de Farias Siqueira

**APLICAÇÃO DE MODELOS MISTOS E MODELOS LINEARES  
GENERALIZADOS MISTOS NA ESTIMAÇÃO DE COMPONENTES DE  
VARIÂNCIA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Prof. Dr. Tiago Almeida de Oliveira

Campina Grande  
Fevereiro de 2014

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S618a Siqueira, Evelyne de Farias.

Aplicação de modelos mistos e modelos lineares generalizados mistos na estimação de componentes de variância [manuscrito] / Evelyne de Farias Siqueira. - 2014.

42 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Componentes de Variância. 2. Herdabilidade. 3. Valores Genéticos. I. Título.

21. ed. CDD 576.5021

Evelyne de Farias Siqueira

**APLICAÇÃO DE MODELOS MISTOS E MODELOS LINEARES  
GENERALIZADOS MISTOS NA ESTIMAÇÃO DE COMPONENTES DE  
VARIÂNCIA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 27/02/2014

**Banca Examinadora:**



---

Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba -  
DE/CCT  
Orientador



---

Prof. Dr. Gustavo Henrique Esteves  
Universidade Estadual da Paraíba -  
DE/CCT  
Examinador



---

Prof. Msc. Sílvio Fernando Alves Xavier  
Júnior  
Universidade Estadual da Paraíba -  
DE/CCT  
Examinador

Em memória de minha querida mãe Inácia Lucena Meira.

MINHA HOMENAGEM

A meu esposo Francisco Medeiros, por me possibilitar,  
com apoio e amor, concluir este trabalho.

E a toda minha família.  
OFEREÇO E DEDICO.

# Agradecimentos

Agradeço de forma muito especial...

A DEUS, pela força para a realização desse trabalho;

Ao meu pai, Manoel Marcos de Farias Meira pelo incentivo e por tudo que me ensinou;

Aos meus irmãos, Emília e Emanuel pela amizade, confiança e grandes incentivos;

As minhas tias, Maria de Lourdes e Josefa pelos conselhos sábios;

Ao meu sobrinho Guilherme, por me proporcionar tantos momentos de alegria e descontração;

A minha cunhada Márcia, pelos constantes incentivos, carinho e amizade;

A todos os professores de minha graduação, que de forma direta ou indireta contribuíram para que o presente trabalho fosse realizado com sucesso;

Ao professor Tiago Almeida de Oliveira, pela amizade, pelos ensinamentos e pela orientação à elaboração deste trabalho;

Aos membros da banca examinadora, professor Gustavo Henrique Esteves e professor Sílvio Fernando Alves Xavier Júnior, pela disponibilidade e pelas sugestões para o enriquecimento deste trabalho;

Aos meus colegas de turma, em especial as amigas Juliana, Moniclaudia e Socorro, por ajudar a tornar a vida acadêmica muito mais divertida. Obrigada pelos momentos de risos, conversas e pela sincera e sólida amizade.

Ao meu esposo Francisco Medeiros, por compreender a importância dessa conquista e aceitar a minha ausência quando necessário.

A todos que foram de alguma forma responsáveis por essa vitória tão esperada ao longo desses cinco anos, meus sinceros agradecimentos.

“A essência do conhecimento consiste  
em aplicá-lo, uma vez possuído”.

- Confúcio.

# Resumo

No contexto da genética quantitativa, a estimação de componentes de variância e a predição dos valores genéticos são procedimentos frequentes no melhoramento de características de importância econômica. A teoria dos modelos lineares mistos, aplicada dentro do contexto do melhoramento genético animal, tem um importante desenvolvimento e uma ampla utilização nos programas de melhoramento e na pesquisa aplicada, o que tem possibilitado a avaliação de genótipos em diferentes programas de seleção e o melhoramento de espécies de interesse econômico em uma grande variedade de condições ambientais. Em consequência das diferenças genéticas da população, de ambiente, do tipo de análise e do método de estimação de componentes de variância, estas estimativas podem variar consideravelmente. O objetivo deste trabalho é estimar componentes de variância pelo método da Máxima Verossimilhança Restrita através dos modelos mistos para a característica quantitativa e modelos lineares generalizados mistos para a característica binária, e prever valores genéticos para os animais sem registro fenotípico no banco de dados simulados através do *software* de programação estatística R. Para as características quantitativa e binária em estudo o uso do método da Máxima Verossimilhança Restrita por modelos lineares mistos e modelos lineares generalizados mistos foram adequados para estimar parâmetros genéticos. As herdabilidades encontradas foram alta para a característica quantitativa e moderada para a característica binária.

**Palavras-chave:** Componentes de variância, herdabilidade, valores genéticos.



# Abstract

In the context of quantitative genetics, the estimation of variance components and the prediction of genetic values are frequent procedures in improvement characteristics of economic importance. The theory of linear mixed models applied within the context of the animal genetic improvement, has had an important development and wide use in breeding programs and applied research, which has enabled the evaluation of genotypes in different programs of selection and species improvement of economic interest in a wide variety of environmental conditions. In consequence of the population genetics differences, environment, type of analyses and the method of estimation variance components, these estimates may vary significantly. The aim of this study is to estimate variance components by the method of Restricted Maximum Likelihood through mixed model for quantitative trait and generalized linear mixed models for binary characteristic, and to predict breeding values for animals without phenotypic record in the database simulated through the software of statistical programming R. For quantitative and binary features under study, we use the method of Restricted Maximum Likelihood by mixed models and generalized linear mixed models were adequate to estimate genetic parameters. Heritability estimates obtained were high for the quantitative characteristic and moderate for binary characteristic.

**Keywords:** Variance components, heritability, genetic values.

# Sumário

## Lista de Tabelas

## Lista de abreviaturas

<b>1</b>	<b>Introdução</b>	p. 13
<b>2</b>	<b>Objetivos</b>	p. 15
2.1	Geral . . . . .	p. 15
2.1.1	Específicos . . . . .	p. 15
<b>3</b>	<b>Revisão de Literatura</b>	p. 16
3.1	Caprinovinocultura . . . . .	p. 16
3.2	Componentes de Variância . . . . .	p. 17
3.2.1	Componentes de Variância Genética . . . . .	p. 17
3.3	Coeficiente de Herdabilidade . . . . .	p. 18
3.4	Estimação dos parâmetros . . . . .	p. 19
3.4.1	Método da Máxima Verossimilhança Restrita . . . . .	p. 19
3.5	Modelos Lineares Mistos . . . . .	p. 20
3.5.1	Definição do Modelo . . . . .	p. 21
3.5.1.1	Modelo Animal . . . . .	p. 23
3.6	Modelo Linear Generalizado . . . . .	p. 24
3.6.1	Família Exponencial . . . . .	p. 25
3.6.2	Modelo Binomial . . . . .	p. 25

3.7	Modelo Linear Generalizado Misto . . . . .	p. 26
3.7.1	Preditor Linear $\eta$ . . . . .	p. 27
3.7.1.1	Função de Ligação <i>probit</i> . . . . .	p. 27
4	<b>Material e Métodos</b>	p. 29
5	<b>Resultados e Discussão</b>	p. 31
6	<b>Conclusão</b>	p. 36
7	<b>Referências</b>	p. 37
	<b>Apêndice A – Programa em linguagem R utilizado para a análise</b>	p. 40

# Lista de Tabelas

1	Conjunto de dados com a característica quantitativa ( $Q$ ) e a característica binária ( $B$ ) . . . . .	p. 30
2	Observações referentes a genealogia dos animais com informações do pai, mãe e sexo. . . . .	p. 30
3	Estimativa dos componentes de variância para a característica quantitativa ( $Q$ ). . . . .	p. 31
4	Estimativa dos componentes de variância para a característica binária ( $B$ ). . . . .	p. 32
5	Valores genéticos e códigos dos animais em relação ao top 10 de todos os animais avaliados para a característica quantitativa ( $Q$ ). . . . .	p. 33
6	Valores genéticos e códigos dos animais em relação ao top 10 de todos os animais avaliados para a característica binária ( $B$ ). . . . .	p. 34
7	Melhores animais dentre os top 10 para as características quantitativa ( $Q$ ) e binária ( $B$ ) analisadas. . . . .	p. 35

# Lista de abreviaturas

**BLUP:** *best linear unbiased predictor*

**EMM:** equações de modelos mistos

**fdp:** função de densidade de probabilidade

**MLG:** modelo linear generalizado

**MLGM:** modelo linear generalizado misto

**REML:** *Restricted Maximum Likelihood.*

# 1 Introdução

Estudos práticos em diversas áreas do conhecimento envolvem o uso de algum modelo de regressão, com o intuito de descrever comportamentos e explicar certos fenômenos. O modelo linear misto vem sendo usado frequentemente, mas quase sempre utilizando a suposição de que as observações são independentes. Apesar disso, é comum nas áreas agrônômica e biológica dentre outras, situações em que os dados apresentam comportamentos dependentes entre si.

A metodologia de modelos mistos (modelos que incluem simultaneamente efeitos fixos e efeitos aleatórios) permite uma estimação eficiente de componentes de variância e valores genéticos com dados provenientes de relações familiares genéticas complexas, tamanhos de família variáveis, e acasalamentos preferenciais.

Atualmente, o modelo misto é comumente utilizado em várias áreas da ciência como uma ferramenta muito flexível para a análise de dados nos quais as observações estão agrupadas por fatores de efeitos aleatórios, de maneira que existe uma dependência entre observações em cada grupo (ROSA, 2013).

Segundo Cruz (2012), componentes de variância são as variâncias associadas aos efeitos aleatórios de um modelo estatístico, sendo que seu conhecimento é de vasta importância no melhoramento genético, pois o método de melhoramento e a população a serem utilizadas dependem de algumas informações que podem ser obtidas a partir desses componentes. Dessa forma a obtenção dos componentes de variância tem sido de grande interesse no melhoramento genético, sendo possível estimar a variância genotípica a partir dos dados fenotípicos observados.

O estudo de métodos de estimação de componentes de variância tem sido cada vez mais importante na cultura de caprinos e ovinos no Brasil, uma vez que a busca do valor verdadeiro do componente permitirá correta predição dos valores genéticos. Portanto, para obtenção dos valores genéticos, os componentes de variância precisam ser estimados de forma acurada para que as diferenças entre o valor predito e o valor verdadeiro sejam

minimizadas.

Para a predição dos valores genéticos, é necessário o conhecimento das variâncias genéticas aditivas e variâncias fenotípicas da característica avaliada ou de sua herdabilidade (CARNEIRO JÚNIOR et al, 2004). Essas variâncias são desconhecidas de modo frequente tornando necessário estimar primeiro componentes de variância e, assim, prever valores genéticos num mesmo conjunto de dados (KENNEDY, 1981).

A técnica de modelos lineares generalizados (MLG), desenvolvida por Nelder e Wedderburn (1972), permite a generalização dos modelos lineares clássicos de variáveis contínuas, de maneira que toda a estrutura para a estimação em modelos lineares normais pode ser estendida para os modelos não lineares. Onde os modelos clássicos são casos especiais de modelos lineares generalizados (RESENDE e BIELE, 2002).

Um modelo linear generalizado é definido por: (i) um componente aleatório, representado pelas variáveis aleatórias independentes, pertencentes a família exponencial; (ii) um componente sistemático denominado preditor linear, no qual as variáveis explicativas são representadas como uma soma linear de seus efeitos; (iii) uma função de ligação que relaciona o componente aleatório ao sistemático, isto é, associando a média ao preditor linear.

Os modelos lineares generalizados mistos (MLGM) permitem modelar a probabilidade de características assumirem determinados valores, por meio da utilização de funções de ligação, em que esta função une o componente aleatório e o componente sistemático. No melhoramento animal, a função de ligação mais utilizada é a *probit*, conhecida como modelo *threshold*, mas existem outras funções que podem ser exploradas e amplamente utilizadas em outras áreas (GARCIA et al., 2012).

Diversos métodos de estimação dos componentes de variância têm sido utilizados, mas o método de escolha pelos programas de melhoramento genético no Brasil atualmente é o método da Máxima Verossimilhança Restrita (FARIA et al., 2008).

O método REML, proposto por PATTERSON e THOMPSON (1971) surgiu a partir de esforços na obtenção de melhores estimadores de componentes de variância, é o mais recomendado para dados desbalanceados, sendo assim o mais adequado para estimação de parâmetros genéticos de dados de melhoramento animal. Logo esse método evita erros de pequenas amostras quando associadas aos efeitos fixos e dados originários de populações submetidas à seleção. O REML tem importante propriedade na estimação de parâmetros da população livre de viés devido à seleção (YAMAKI, 2006).

## 2 Objetivos

### 2.1 Geral

Este trabalho tem por objetivo fazer uma aplicação de modelos mistos e modelos lineares generalizados mistos para estimar componentes de variância mediante dados simulados.

#### 2.1.1 Específicos

- Aplicar o modelo linear misto na estimação de componentes de variância para a característica quantitativa;
- Aplicar o modelo linear generalizado misto na estimação de componentes de variância para a característica binária;
- Utilizar o método de estimação da Máxima Verossimilhança Restrita (REML), nos modelos lineares mistos e modelos lineares generalizados mistos, por meio de dados simulados;
- Obter as estimativas de herdabilidade para as características quantitativa e binária mediante dados simulados;
- Predizer os valores genéticos dos animais no modelo linear misto e modelo linear generalizado misto;
- Classificar os melhores animais para as características quantitativa e binária.



## 3 Revisão de Literatura

### 3.1 Caprinovinocultura

A Caprinovinocultura é uma atividade econômica explorada em todos os continentes do mundo, sendo exercida em áreas sob diversas características climáticas. Entretanto, apenas em alguns países essa atividade apresenta expressão econômica, muitas vezes, desenvolvida de forma empírica e ampla, com baixos níveis de tecnologia (ROSANOVA, 2004).

Na região Nordeste do Brasil, a caprinovinocultura é considerada uma atividade de subsistência normalmente com produtividade baixa e desenvolvida por produtores desprovidos de capital financeiro e de recursos tecnológicos . Atualmente a produção desses animais vem se caracterizando como atividade de grande importância econômica, social e cultural, desempenhando um papel fundamental no desenvolvimento do Nordeste (COSTA et al, 2008).

O Brasil possui um rebanho efetivo de 9,38 milhões de caprinos e 17,3 milhões de ovinos que, somados, representam 26,68 milhões de cabeças. O efetivo de caprino no Nordeste representa 90,98% de todo o rebanho nacional, enquanto o rebanho ovino 57,24%. Os principais estados detentores dos maiores rebanhos de caprinos e ovinos, respectivamente, são: Bahia, com 35,72% e 24,64%; Piauí, com 18,00% e 11,21%; Pernambuco, com 25,09% e 14,89% e Ceará, com 13,61% e 17,18% <sup>1</sup>

A região Nordeste apresenta uma tradição na criação de caprinos (bodes e cabras) e ovinos (carneiros e ovelhas) , principalmente pela produção de carne e leite de ótima qualidade, por um custo baixo, sendo atualmente um componente importante no sustento da família para populações rurais de baixa renda.

Com os novos rumos do comércio, e o aumento da competitividade, tende a buscar

---

<sup>1</sup><http://www.farmpoint.com.br/cadeia-produtiva/especiais/analise-do-desenvolvimento-do-rebanho-ovino-e-caprino-no-brasil-em-2011-81339n.aspx>

saídas para um crescimento em quantidade e qualidade na produção de animais geneticamente superiores. O surgimento de novas técnicas de reprodução através do melhoramento genético animal, tem proporcionado um aumento de produtividade e melhoria de várias características de interesse. Desta forma, o melhoramento genético é considerado fundamental e importante por apresentar resultados de longo prazo e tem como objetivo principal utilizar a variação genética para melhorar a produção de animais nas gerações seguintes (ROGES et al., 2010).

De acordo com Cruz (2012), o melhoramento genético é um meio a alcançar aumentos na produtividade e qualidade e possui, em relação a técnicas de natureza ambiental, a vantagem de promover alterações hereditárias, ou seja, de ser capaz de transmitir as boas características, alcançadas pelo melhoramento, aos seus descendentes.

Em estatística genética, os parâmetros genéticos são obtidos a partir de várias metodologias de análise estatística que vem sendo utilizadas como ferramentas a fim de minimizar a interferência de fatores ambientais nos resultados da correta estimação de componentes de variância e predição de valores genéticos.

## **3.2 Componentes de Variância**

Segundo Falconer (1987), do ponto de vista genético, o estudo de um caráter quantitativo baseia-se na sua variação total, que pode ser atribuída a diversos fatores. A grandeza dos componentes de variância determina as propriedades genéticas da população para um determinado caráter. Já do ponto de vista estatístico, componentes de variância, são variâncias associadas aos efeitos aleatórios de um modelo matemático (BARBIN, 1993).

Componentes de variâncias são as variâncias associadas aos efeitos aleatórios de um modelo estatístico, que podem ser estimados através de diferentes métodos. A obtenção desses componentes é de grande interesse em melhoramento genético, logo é possível estimar a variância genotípica a partir de dados fenotípicos observados (CRUZ, 2012).

### **3.2.1 Componentes de Variância Genética**

O valor observado, em um indivíduo, é chamado de valor fenotípico, onde pode ser decomposto devido às influências do genótipo e do ambiente no qual ele se desenvolveu. Compreende-se por genótipo a combinação de genes do indivíduo, e por ambiente qualquer outro fator que possa influenciar o fenótipo (FALCONER, 1987).

No estudo da variação de caracteres quantitativos adota-se o modelo básico  $F = G + A$ , em que (F) é o valor fenotípico, estimado a partir de dados mensurados dos indivíduos, resultante da ação do genótipo (G), sob influência do ambiente (A). Analogamente, a variância fenotípica (VF) é composta pela variância genotípica (VG ou  $\sigma^2_G$ ) e pela variância atribuída aos desvios pelo ambiente (VA).

De acordo com Cruz (2012), a variância genotípica é estabelecida por três outros componentes, conforme descrito a seguir:

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$$

em que:

$\sigma^2_A$  – **Variância genética aditiva:** é o componente mais importante, pois é a principal causa de semelhança entre parentes. Além disso, é o único componente que pode ser estimado pelas observações feitas na população;

$\sigma^2_D$  – **Variância genética de dominância:** é a variância atribuída aos desvios da dominância ou proporcionada pelas interações intra-alélicas (ou entre alelos de um mesmo loco);

$\sigma^2_I$  – **Variância genética epistática:** é a variância atribuída aos efeitos epistáticos resultantes de interações inter-alélicas (ou entre alelos de diferentes locos).

Neste trabalho todo o desenvolvimento foi com base em um modelo que contempla apenas efeitos de variância genética aditiva, já que é o que interessa no melhoramento, pois é o componente que transmite à sua descendência.

### 3.3 Coeficiente de Herdabilidade

De acordo com Falconer (1987), a herdabilidade é definida como a razão da variância genética aditiva para a variância fenotípica. Onde os valores genotípicos e fenotípicos são considerados como desvios da média da população. Em outros termos, a herdabilidade expressa o grau de confiança do valor fenotípico como um guia para o valor genético, ou o grau de semelhança entre o valor fenotípico e o valor genético.

A herdabilidade permite indicar o potencial para a seleção fenotípica e melhorar geneticamente uma característica. Uma herdabilidade alta sugere que a seleção individual produzirá um melhoramento genético rápido, porém uma herdabilidade baixa indica que o

aumento na seleção será demorado e outros métodos deveriam ser adotados para melhorar a característica (FERREIRA, 2011).

A herdabilidade pode variar de 0 a 1 (0% a 100%). Valores baixos podem significar que grande parte da variação é de natureza ambiental e se os valores são altos significa que grande parte da variação é de natureza genética (VAYEGO, 2007).

## 3.4 Estimação dos parâmetros

Atualmente, na literatura pode-se encontrar diversos métodos para estimação de parâmetros em modelos de regressão e novos métodos são apresentados a cada dia. No modelo linear misto, os métodos mais utilizados são o de Máxima Verossimilhança e o de Máxima Verossimilhança Restrita. Será apresentado nesta seção o método da Máxima Verossimilhança Restrita o qual foi utilizado neste estudo.

### 3.4.1 Método da Máxima Verossimilhança Restrita

O método da Máxima Verossimilhança Restrita (REML), sugerido por Patterson & Thompson (1971), visa estimar componentes de variância, sendo o mais recomendado para dados desbalanceados, e conseqüentemente o mais adequado para estimação de parâmetros genéticos de dados de melhoramento animal. Desta forma esse método evita erros de pequenas amostras quando associadas aos efeitos fixos e dados originários de populações submetidas à seleção. O REML tem importante propriedade na estimação de parâmetros da população livre de viés devido à seleção (YAMAKI, 2006).

Simonato (2009), definiu que no REML as combinações lineares dos dados chamados de incrementos generalizados, são usadas no lugar das observações. Onde esses incrementos filtram as tendências e estima apenas os parâmetros de covariância, conduzindo a um estimador não tendencioso. O método REML supõe a normalidade dos dados, é iterativo e fornece sempre estimativas não negativas dos componentes de variância .

A justificativa de Simonato (2009) por trás da Máxima Verossimilhança Restrita é talvez melhor expressa em um caso muito simples. Suponha que  $X_1, \dots, X_n$  são variáveis aleatórias independentes, cada  $X_i \sim N(\mu, \sigma^2)$  com  $\mu$  e  $\sigma^2$  desconhecidos. Como sabemos, o estimador de máxima verossimilhança de  $\mu$  e  $\sigma^2$  é  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_i X_i$  e  $\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$ , respectivamente. Todavia, esta definição de  $\hat{\sigma}^2$  é um estimador tendencioso, enquanto que o estimador mais usual é  $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$  não é tendencioso. No método da Máxima

Verossimilhança Restrita, as observações são divididas em duas partes independentes, uma referente aos efeitos fixos e outra referente aos efeitos aleatórios, de forma que a função de densidade de probabilidade das observações é dada pela soma das funções de densidade de probabilidade de cada parte. A maximização da função de densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o viés resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo. Assim, o estimador REML leva em conta os graus de liberdade envolvidos nas estimativas dos efeitos fixos do modelo e as equações com dados desbalanceados são não viciados e de variância mínima.

### 3.5 Modelos Lineares Mistos

Um modelo linear misto é aquele que apresenta tanto fatores de efeitos fixos como fatores de efeitos aleatórios, além do erro experimental e da constante  $\mu$  e sua análise de variância apresenta algumas propriedades, como a composição das esperanças matemáticas dos quadrados médios cujo conhecimento admite o estabelecimento correto dos testes de hipóteses. Caso o interesse do pesquisador reside na estimação de componentes de variância, métodos apropriados devem ser utilizados (CAMARINHA FILHO, 2002).

Os modelos mistos vem sendo empregados no melhoramento animal para obter o melhor preditor não viesado (Blup) dos valores genéticos. Contudo esse método supõe que os componentes de variância sejam conhecidos, o que dificilmente ocorre na prática. Logo que esses componentes são desconhecidos, aplica-se a estratégia de substituição dos valores paramétricos dos componentes de variância por estimativas de Máxima Verossimilhança Restrita (REML), recebendo o nome de Blup Empírico (CARNEIRO JÚNIOR et al., 2010).

Em consequência das diferenças genéticas da população, de ambiente, do tipo de análise e do método de estimação de componentes de variância, entre outros, estas estimativas podem variar consideravelmente (DE PAULA, 2006). Neste sentido, a predição dos efeitos aleatórios do modelo misto, por exemplo, a predição dos valores genéticos de animais, depende das estimações dos componentes de variância, os quais têm sido estimados frequentemente pelo método REML, que tem sido comumente preferido (RESENDE, 2002).

Uma razão de se adotar um modelo misto é a possibilidade de se fazer predições de efeitos aleatórios, na presença de efeitos fixos, que são de grande importância em genética

e melhoramento.

### 3.5.1 Definição do Modelo

Segundo Camarinha Filho (2002), o modelo linear misto é dado da seguinte forma:

$$y_{ijk} = \mu + \beta_i + a_j + \varepsilon_{ijk} \quad (3.1)$$

em que,

$y_{ijk}$  é a observação referente à  $k$ -ésima repetição do nível  $i$  de uma fonte de efeitos fixos ao nível  $j$  de uma fonte de efeitos aleatórios;

$\mu$  é uma constante inerente a todas as observações;

$\beta$  é o efeito do nível  $i$  do fator fixo;  $i = 1, \dots, p$ ;

$a$  é o efeito do nível  $j$ , do fator aleatório, no nível  $i$  do fator fixo,  $j = 1, \dots, q$ ;

$\varepsilon_{ijk}$  é o erro aleatório associado a observação  $y_{ijk}$ .

Usando a notação matricial, o modelo linear misto descrito é denotado por:

$$y = X\beta + Za + \varepsilon \quad (3.2)$$

em que,  $y$  é o vetor de observações;  $X$  é a matriz de incidência dos efeitos fixos (conhecida);  $\beta$  é o vetor de efeitos fixos desconhecidos;  $Z$  é a matriz de incidência dos efeitos aleatórios (conhecida);  $a$  é o vetor de efeitos aleatórios desconhecidos;  $\varepsilon$  é o vetor de erros aleatórios.

Uma suposição na versão clássica do modelo, é que tanto os erros como os efeitos aleatórios são normalmente distribuídos e assumidos independentes entre si. Sob essas suposições, inferências para os parâmetros fixos e aleatórios dos modelos é obtido através do programa R.

Dadas as especificações do modelo descritas acima, a distribuição de  $y$ ,  $a$  e  $\varepsilon$  podem ser descritas como:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{X}\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{ZGZ}' + \mathbf{R} & \mathbf{ZG} & \mathbf{R} \\ \mathbf{GZ}' & \mathbf{G} & \emptyset \\ \mathbf{R} & \emptyset & \mathbf{R} \end{bmatrix} \right),$$

em que,  $\emptyset$  é uma matriz de zeros;  $\mathbf{R} = \mathbf{I}\sigma_e^2$  é a matriz de variância e covariância residual,  $\mathbf{I}$  é a matriz identidade de ordem igual ao número de animais em estudo;  $\mathbf{G} = \mathbf{A}\sigma_a^2$  é a

matriz de variância e covariância dos efeitos genéticos aditivos,  $\mathbf{A}$  é a matriz de parentesco genético entre os animais.

De modo geral, assume-se que os vetores  $a$  e  $\varepsilon$  são independentes entre si e distribuídos normalmente com média zero e são variáveis não correlacionadas, com matrizes de variâncias-covariâncias, respectivamente,  $\mathbf{G}$  e  $\mathbf{R}$  matrizes positivas definidas, por hipótese, e, por isso, não singulares, dadas por:

$$\text{Var}(a) = E(aa') = \mathbf{G} \text{ e } \text{Var}(\varepsilon) = E(\varepsilon\varepsilon') = \mathbf{R}.$$

Assim, tem-se que:

$$V = \text{Var}(y) = \text{Var}(X\beta) + \text{Var}(Za) + \text{Var}(\varepsilon) = Z\text{Var}(a)Z' + R = ZGZ' + R.$$

Assume-se, ainda, que  $V$  é não singular, e

$$E(y) = E(X\beta + Za + \varepsilon) = X\beta,$$

deste modo,  $y \sim N(X\beta; ZGZ' + R)$ .

Henderson (1950) desenvolveu as equações de modelos mistos (EMM) para estimar  $\beta$  e  $a$  simultaneamente, sem a necessidade da obtenção da inversa  $[ZGZ' + R]^{-1}$ . As equações de modelos mistos podem ser encontradas a partir da maximização (em termos de  $\beta$  e  $a$ ) da função de densidade de probabilidade conjunta de  $y$  e  $a$ .

A função densidade de probabilidade de  $y$  é dada por:

$$f(y, a) = \frac{1}{(2\pi)^{n/2} [ZGZ' + R]^{1/2}} e^{-\frac{1}{2} [(y - X\beta)' (ZGZ' + R)^{-1} (y - X\beta)]}. \quad (3.3)$$

Conforme o modelo dado em (3.2) e a função densidade de probabilidade de  $y$  dada em (3.3), tem-se que a função densidade de probabilidade conjunta de  $y$  e  $a$  pode ser escrita como o produto entre a função de densidade condicional de  $y$ , dado  $a$ , e a função densidade de probabilidade de  $a$ . Logo,

$$f(y, a) = f(y/a) \cdot f(a).$$

Assim,

$$f(y, a) = \frac{1}{(2\pi)^{n/2}[R]^{1/2}} e^{-\frac{1}{2}[(y-x\beta-Za)'(R)^{-1}(y-x\beta-Za)]} \cdot \frac{1}{(2\pi)^{n/2}[G]^{1/2}} e^{-\frac{1}{2}[(a-0)'(G)^{-1}(a-0)]},$$

sendo  $[R]$  e  $[G]$  os determinantes das matrizes de covariâncias.

Para se efetuar à maximização de  $f(y, a)$ , pode-se utilizar a transformação por logaritmo. Isso é possível, visto que, sendo  $f(y, a)$  e  $\log [f(y, a)]$  funções contínuas e crescentes no espaço  $R^+$ , seus pontos de máximo são coincidentes dentro do espaço de  $[\beta, a]$  e  $ZGZ' + R$ .

Assim, fazendo-se  $L = \log[f(y, a)]$ , tem-se:

$$L = \frac{1}{2}2n \log(2\pi) - \frac{1}{2}(\log R + \log G) - \frac{1}{2}(y'R^{-1}y - 2y'R^{-1}X\beta - 2y'R^{-1}Za + 2\beta'X'R^{-1}Za + \beta'X'R^{-1}X\beta + aZ'R^{-1}Za + a'Ga).$$

Derivando-se  $L$  em relação a  $\beta$  e  $a$ , e tornando-se as derivadas parciais identicamente nulas, obtêm-se:

$$\begin{bmatrix} \frac{\partial L}{\partial \beta} \\ \frac{\partial L}{\partial a} \end{bmatrix} = \begin{bmatrix} -X'R^{-1}y + X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{a} \\ -Z'R^{-1}y + Z'R^{-1}X\hat{\beta} + Z'R^{-1}Z\hat{a} + G^{-1}\hat{a} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

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

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

Essas são as equações de modelos mistos (EMM), que permitem obter estimativas para os efeitos fixos ( $\hat{\beta}$ ) e predições para os efeitos aleatórios ( $\hat{a}$ ).

### 3.5.1.1 Modelo Animal

O modelo utilizado para estimar os componentes de variância é dado por:

$$y_{ij} = \mu_j + animal_{ij} + \varepsilon_{ij} \quad (3.4)$$



onde  $y_{ij}$  é o registro fenotípico de animais  $i$  para característica  $j$ ,  $\mu_j$  é a média geral para a característica  $j$ ,  $animal_{ij}$  é o efeito poligênico aleatório do animal  $i$  para característica  $j$ , e  $\varepsilon_{ij}$  é o efeito aleatório residual para o animal  $i$  na característica  $j$ .

### 3.6 Modelo Linear Generalizado

Os modelos lineares generalizados (MLG) foram introduzidos nos anos 70, tendo um grande impacto no desenvolvimento da estatística em diversas áreas do conhecimento. A princípio seu uso era restrito a um grupo de investigadores, pela falta de bibliografia acessível e à complexidade inicial do primeiro *software* dirigido à sua aplicação. Levaram cerca de 20 anos para que a teoria e prática dos modelos lineares generalizados passasse para domínio público, e este avanço só foi possível a partir do momento que o *software* existente passou a ser melhor entendido. Pode dizer-se que, o conhecimento adequado dessa metodologia é imprescindível para qualquer pessoa que utilize métodos estatísticos (TURKMAN e SILVA, 2000).

A metodologia MLG permite que a distribuição de probabilidade da variável resposta seja qualquer distribuição que pertença à família exponencial e utiliza métodos análogos do modelo linear com dados de distribuição normal (NELDER e WEDDERBURN, 1972).

Segundo Cordeiro e Demétrio (2009), um modelo linear generalizado é definido por uma distribuição de probabilidade pertencente a família exponencial para a variável resposta, um conjunto de variáveis explicativas descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável resposta e a estrutura linear. Dentre os métodos estatísticos para a análise de dados que são casos especiais de modelos lineares generalizados há o modelo *probit* para o estudo de proporções, englobando a distribuição binomial.

Em resumo, são fundamentais na definição de um modelo linear generalizado:

- (i) definição da distribuição da variável resposta;
- (ii) definição do preditor linear;
- (iii) definição da função de ligação.

### 3.6.1 Família Exponencial

Os modelos lineares generalizados pressupõem que a variável resposta tenha uma distribuição pertencente a família exponencial. Então, diz-se que uma variável aleatória  $Y$  tem distribuição pertencente a família exponencial se a sua função de densidade de probabilidade (f.d.p.) puder ser escrita na forma

$$f(y|\theta, \phi) = \exp\{y\theta - b(\theta)/a(\phi) + c(y, \phi)\}, \quad (3.5)$$

onde  $\theta$  e  $\phi$  são parâmetros escalares,  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções reais conhecidas, sendo  $E[Y] = \mu = b'(\theta)$  e  $\text{Var}(Y) = a(\phi)b''(\theta) = a(\phi)V(\mu)$ .

Em que  $\theta$  é a forma canônica do parâmetro de localização e  $\phi$  é um parâmetro de dispersão suposto, geralmente conhecido. Neste caso, a distribuição descrita em (3.5) faz parte da família exponencial uniparamétrica. No entanto, quando o parâmetro  $\phi$  é desconhecido a distribuição pode ou não fazer parte da família exponencial bi-paramétrica. Admite-se, ainda, que a função  $b(\cdot)$  é diferenciável e que o suporte da distribuição não depende dos parâmetros.

### 3.6.2 Modelo Binomial

O modelo binomial descreve o número de ocorrências de uma variável binária em um certo número de tentativas, em função de variáveis preditoras contínuas ou discretas. Sendo úteis principalmente quando temos respostas binárias e proporções.

Dados onde a variável resposta possui natureza dicotômica, isto é, assume somente dois valores, geralmente 0 para não ocorrência do fenômeno e 1 para a ocorrência, pode ser considerada como provenientes de uma distribuição Bernoulli, com probabilidades  $p$  (para 1) e  $1-p$  (para 0).

Suponha que se  $Y$  for tal que  $mY$  tem uma distribuição binomial com parâmetros  $m$  e  $\pi$  ( $Y \sim B(m, \pi)/m$ ), a sua f.d.p. é dada por:

$$\begin{aligned} f(y|\pi) &= \binom{m}{ym} \pi^{ym} (1-\pi)^{m-ym} \\ &= \exp \left[ ym \ln \pi + m(1-y) \ln(1-\pi) + \ln \binom{m}{ym} \right] \\ &= \exp \left[ m(y\theta - \ln(1+e^\theta)) + \ln \binom{m}{ym} \right]. \end{aligned}$$

Desta forma, o componente aleatório do modelo linear generalizado misto para a variável binária, tem uma distribuição que pertence a família exponencial, e para o ajuste do modelo, utiliza-se a função de ligação *probit*, considerando a distribuição binomial, para estimar a probabilidade da característica binária ( $B$ ) assumir o valor 0 ou 1.

### 3.7 Modelo Linear Generalizado Misto

De acordo com Resende e Biele (2002), a técnica de modelos lineares generalizados (MLG), desenvolvida por Nelder e Wedderburn (1972), permite a generalização dos modelos lineares clássicos de variáveis contínuas, de maneira que toda a estrutura para a estimação em modelos lineares normais pode se estender para os modelos não lineares, onde os modelos clássicos são casos especiais de modelos lineares generalizados.

Segundo Cordeiro e Demétrio (2009), os modelos lineares generalizados podem ser vistos como uma unificação dos modelos de regressão linear e não linear, o qual incorpora uma ampla família de distribuições da variável resposta. Esses modelos envolvem uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de observações independentes, sendo que:

- i) a variável resposta, componente aleatório do modelo, tem uma distribuição pertencente à família exponencial de distribuições;
- ii) as variáveis explanatórias entram na forma de uma estrutura linear, constituindo o componente sistemático do modelo;
- iii) a ligação entre o componente aleatório e o sistemático é feita através de uma função adequada, denominada função de ligação.

No contexto dos modelos mistos com efeitos fixos e aleatórios, o componente sistemático dos modelos lineares generalizados mistos refere-se aos próprios modelos lineares mistos clássicos exceto o termo do erro, sendo que, para o caso de variáveis binomiais, a função de ligação deve ser incorporada aos preditores (melhores preditores lineares não viciados – BLUP) dos efeitos aleatórios e estimadores (de Máxima Verossimilhança Restrita – REML) dos componentes de variância dos efeitos aleatórios (RESENDE e BIELE, 2002).

### 3.7.1 Preditor Linear $\eta$

Através da teoria dos modelos lineares generalizados mistos, os efeitos fixos e aleatórios, que influenciam determinada característica, são combinados formando um preditor linear, descrito por:

$$\eta = X\beta + Za \quad (3.6)$$

em que,

$\eta$  é o preditor linear;

$\beta$  é o vetor de efeitos fixos;

$a$  é o vetor dos valores genéticos aditivos;

$X$  e  $Z$  são, respectivamente, as matrizes de incidência dos efeitos  $\beta$  e  $a$ .

Assim como um modelo linear misto o modelo para o vetor de observações  $y$  é obtido pela adição de um vetor de resíduos,  $\varepsilon \sim N(0, R)$  como segue,

$$y = \eta + \varepsilon = X\beta + Za + \varepsilon.$$

Posteriormente, de acordo com a distribuição de probabilidade da característica analisada é definida a função de ligação apropriada. Neste caso, como a variável binária segue uma distribuição binomial, a função *probit* (modelo *threshold*) é utilizada. Quando a suposição de normalidade pode ser assumida, a função de ligação identidade é adotada, com isso, utiliza-se o modelo clássico para estimação de parâmetros genéticos e predição dos valores genéticos dos animais. Dessa forma, é importante ressaltar que o modelo *probit* é um caso particular dos modelos lineares generalizados.

#### 3.7.1.1 Função de Ligação *probit*

Quando em um estudo as variáveis explicativas possuem natureza binária, ou seja, a presença ou ausência de determinados atributos, a variável dependente  $Y$  poderá assumir apenas somente um de dois possíveis valores, chamados de  $Y = 1$  ou  $Y = 0$ , dessa forma, é possível calcular as probabilidades como sendo, a probabilidade de sucesso e fracasso.

Através da utilização da teoria dos MLGM é possível modelar a probabilidade da variável estudada, por exemplo, a probabilidade de um animal  $i$  sobreviver à desmama é codificada com o valor 1, ou com o valor 0 em caso contrário.

Considerando que  $\Pi_i = P[Y=1]$  é a probabilidade que a característica assumira o valor

1, a função de ligação *probit*,  $g(\cdot)$ , é dada por:

$$g(\Pi_i) = \Phi^{-1}(\Pi_i)$$

em que:  $\Phi^{-1}$  é a função inversa da normal padrão acumulada.

Na variável binária a fonte de variação não explicada pelo modelo na predição da expressão fenotípica, normalmente denominada de resíduo, está inclusa na aleatoriedade dos valores fenotípicos gerados para o animal  $i$  pela distribuição binomial com probabilidade  $\Pi_i$ .

Conforme a teoria de modelos lineares generalizados mistos, ao se ligar o preditor linear ( $\eta$ ) à função de ligação, é possível obter as probabilidades ( $\Pi_i$ ) de a característica assumir 0 ou 1.

As probabilidades são dadas por:

$$g(\Pi_i) = \Phi^{-1}(\Pi_i) = \eta_i = x'_i\beta + a_i$$

$$\Phi^{-1}(\Pi_i) = x'_i\beta + a_i$$

$$\Pi_i = \Phi(x'_i\beta + a_i).$$

O modelo linear generalizado misto, obtido pela associação do modelo binomial para as respostas, com a função de ligação *probit* conduz ao *modelo de regressão probit*.

De acordo com a teoria dos modelos lineares generalizados mistos os valores para os componentes de variância residual são fixos (RESENDE e BIELE, 2002). Quando utilizamos a função de ligação *probit*, a variância residual  $\hat{\sigma}_e^2$  é desconhecida e, geralmente, é assumida como sendo 1. Este valor é utilizado com o objetivo de tornar a função de verossimilhança identificável (GARCIA et al., 2012).

## 4 Material e Métodos

Os dados utilizados para este estudo foram provenientes do 14<sup>o</sup> Workshop QTL-MAS 2010 em Poznan(Polônia)<sup>1</sup>. Como no banco de dados algumas das informações não estavam explícitas, isto é, de que tipo de indivíduos jovens se tratavam, mas pelas características descritas dos dados e através do peso médio foi considerado de forma didática que a característica  $Q$  possa estar relacionada ao peso adulto ou peso ao sobreano de caprinos ou ovinos, e a característica  $B$  esteja relacionada a precocidade à reprodução, sobrevivência à desmama de caprinos ou ovinos, e entre outras características que tenham a peculiaridade de representarem presença ou ausência. Os dados são analisados de acordo com o trabalho de Szydlowski e Paczynska (2011), na qual foram analisadas duas características. Característica  $Q$  é uma característica quantitativa, e a característica  $B$  é uma característica binária. O banco de dados, é composto por informações de indivíduos jovens (Geração F4: indivíduos 2327 à 3226) onde não tem valor fenotípico observado. O *pedigree* é composto por 3226 indivíduos em 5 gerações (F0 - F4). Há 20 fundadores: 5 machos e 15 fêmeas. Um arquivo *pedigree* está definido na seguinte forma: ID (indivíduo); Parente Macho (pai), Parente Fêmea (mãe), 0 = pais ausentes; Sexo (M = masculino, F= feminino).

Por considerar a relação de parentesco existentes entre os animais utilizou-se o modelo animal descrito em (3.4) para a característica quantitativa, de indivíduos jovens (Geração F4: no qual os indivíduos 2327 à 3226) na qual não tem mensuração. Os dados foram analisados no programa R (R DEVELOPMENT CORE TEAM, 2013), utilizando-se o pacote *pedigreemm* (VAZQUEZ et al., 2010). Este pacote utiliza no processo de estimação dos parâmetros o método da Máxima Verossimilhança Restrita - REML (GARCIA et al., 2012). Para a característica binária foi utilizado o modelo (3.4) também utilizando o pacote *pedigreemm* (VAZQUEZ et al., 2010).

---

<sup>1</sup>Disponível em: <http://jay.up.poznan.pl/qtlmas2010/dataset.html>

Na Tabela 1 temos uma amostra dos dados utilizados neste trabalho.

Tabela 1: Conjunto de dados com a característica quantitativa ( $Q$ ) e a característica binária ( $B$ )

Obs	Animal	Característica $Q$	Característica $B$
1	1	64,22	0
2	2	88,88	1
3	3	64,02	0
4	4	68,51	0
5	5	61,08	0
6	6	70,34	0
7	7	68,57	0
8	8	76,43	0
9	9	73,67	1
10	10	71,49	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
2326	2326	78, 42	0

Na Tabela 2 tem-se os dados referentes a genealogia dos animais (parentesco), com os quais será montada a matriz de parentesco. Os dados foram analisados segundo a metodologia de modelos mistos, considerando-se um modelo animal misto para a característica quantitativa e um modelo animal generalizado misto para a característica binária.

Tabela 2: Observações referentes a genealogia dos animais com informações do pai, mãe e sexo.

Obs	animal	pai	mãe	sexo
1	1	0	0	M
2	2	0	0	M
3	3	0	0	M
4	4	0	0	M
5	5	0	0	M
6	6	0	0	F
7	7	0	0	F
8	8	0	0	F
9	9	0	0	F
10	10	0	0	F
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
3226	3226	1590	2057	M

## 5 Resultados e Discussão

Os resultados foram obtidos por meio do método de estimação da Máxima Verossimilhança Restrita (REML), utilizando o modelo misto unicaracterístico para dados de característica quantitativa ( $Q$ ) e com base em um modelo linear generalizado misto para dados de característica binária ( $B$ ), obtidos de um banco de dados através da internet simulados em um workshop, as estimativas dos componentes de variância de animais jovens (Geração F4: no qual os indivíduos 2327 à 3226) não tem mensuração.

As estimativas obtidas dos componentes de variância para a característica quantitativa ( $Q$ ) analisada é apresentada na Tabela 3.

Tabela 3: Estimativa dos componentes de variância para a característica quantitativa ( $Q$ ).

	Componentes	Variância	Desvio Padrão
Efeitos Aleatórios	Animal	56,810	7,5373
	Residual	47,695	6,9061
Efeitos fixos	Efeito	Estimativa	Desvio Padrão
	Intercepto	69,464	1,823

Os componentes de variância estimados obtidos através da metodologia REML por equações de modelos mistos foram utilizados para calcular a herdabilidade. A estimativa do coeficiente de herdabilidade para a característica quantitativa analisada foi:

$$\hat{h}^2 = \frac{56,810}{56,810 + 47,695} = 0,5436$$

em que,  $\hat{\sigma}_a^2 = 56,810$  é a estimativa da variância genética animal e  $\hat{\sigma}_e^2 = 47,695$  a estimativa da variância genética residual.

Esse valor pode ser considerado como sendo de alta magnitude, e assemelha-se ao apresentado por Carneiro Júnior et al. (2004) em situação semelhante, em que os valores apresentados foram  $\hat{h}^2 = 0,60$  e  $\hat{h}^2 = 0,30$ . Através da estimativa de herdabilidade para a característica de interesse é possível efetuar a seleção dos animais baseando-se em seus



valores genéticos.

De acordo com Pereira (2008), valores de herdabilidade de 0 a 0,15 são considerados baixos; entre 0,15 a 0,30 moderados e acima de 0,30 alta herdabilidade. Os valores estimados de herdabilidade neste trabalho apresentaram-se de moderado a alto, para as características quantitativa e binária respectivamente analisadas, indicando que ao se considerar estas características na seleção de caprinos e ovinos haverá ganho genético no rebanho.

O método REML apresenta estimativas de componentes de variância genética aditiva mais próxima do valor verdadeiro, devido ao fato de que o REML, quando utilizado no modelo animal aproveita todas as informações que os dados contêm, devido ao uso da matriz de parentesco (CARNEIRO JÚNIOR et al., 2004).

As estimativas obtidas dos componentes de variância para a característica binária ( $B$ ) analisada é apresentada na Tabela 4.

Tabela 4: Estimativa dos componentes de variância para a característica binária ( $B$ ).

Efeitos Aleatórios	Componentes	Variância	Desvio Padrão
	Animal	0,3142	0,5605
Residual	1	1	
Efeitos fixos	Efeito	Estimativa	Desvio Padrão
	Intercepto	-0,7213	0,1506

A estimativa do coeficiente de herdabilidade para a característica binária analisada foi:

$$\hat{h}^2 = \frac{0,3142}{0,3142 + 1} = 0,2391$$

em que, no MLGM  $\hat{\sigma}_e^2 = 1$  a estimativa da variância genética residual e  $\hat{\sigma}_a^2 = 0,3142$  a estimativa da variância genética animal.

O componente de variância genético animal foi estimado através da metodologia REML por modelos lineares generalizados mistos. O valor de  $\hat{h}^2 \cong 0,24$ , pode ser considerado de magnitude moderada.

Em um estudo feito por Carneiro Júnior et al. (2004) são utilizados dados simulados na comparação de vários métodos de estimação para componentes de variância de dados animais. Conclui-se que o método REML, pode ser considerado o mais apropriado para estimar componentes de variância para características de baixa herdabilidade em modelos animais.

Segundo Torres Filho et al. (2005), as estimativas de herdabilidade para características reprodutivas são normalmente baixas, devido ao fato de o caráter ser bastante influenciado pelo ambiente. Todavia, Carneiro Júnior et al. (2004) observaram que para herdabilidades menores, a maior parte das diferenças entre os indivíduos não pode ser verificada pelo valor fenotípico.

Calus et al. (2011) utilizando um conjunto de dados simulados através de um modelo animal, encontraram valores próximos aos observados neste trabalho da estimativa do coeficiente de herdabilidade, tanto para característica quantitativa e binária, e foram respectivamente  $\hat{h}^2 = 0,53$  e  $\hat{h}^2 = 0,22$ .

Garcia et al (2012), utilizando modelos com a função de ligação *probit* nas estimativas de herdabilidade encontraram valores considerados de alta herdabilidade, onde esses valores foram  $\hat{h}^2 = 0,47$  e  $\hat{h}^2 = 0,64$ , sendo considerados valores superiores aos encontrados neste trabalho para a característica binária, utilizando a mesma função de ligação.

Por meio dos valores genéticos é possível classificar os melhores animais para serem os melhoradores das gerações seguintes, deste modo foram selecionados os 10 animais com maiores valores genéticos em cada uma das características avaliadas (top 10). Os valores genéticos relacionados a característica quantitativa ( $Q$ ) encontram-se na Tabela 5 .

Tabela 5: Valores genéticos e códigos dos animais em relação ao top 10 de todos os animais avaliados para a característica quantitativa ( $Q$ ).

Animal (código)	Valor Genético	Posição
2	28,99	1°
280	22,35	2°
188	20,65	3°
17	18,55	4°
649	17,69	5°
299	16,29	6°
20	15,98	7°
15	14,36	8°
8	13,78	9°
342	12,34	10°

Os resultados apresentados na Tabela 5 poderiam ser usados pelos criadores com o objetivo de atender à demanda do seu mercado consumidor, criando condições para o aumento da lucratividade e sustentação do rebanho.

Os valores genéticos relacionados a característica binária ( $B$ ) encontram-se na Tabela 6.

Tabela 6: Valores genéticos e códigos dos animais em relação ao top 10 de todos os animais avaliados para a característica binária ( $B$ ).

Animal (código)	Valor Genético	Posição
2	2,63	1°
3	2,32	2°
17	2,16	3°
13	1,96	4°
88	1,25	5°
280	1,21	6°
188	1,19	7°
609	1,09	8°
4	1,05	9°
116	0,97	10°

A prioridade de um criador que trabalha com a fase de cria é a fertilidade. O planejamento para fertilidade garantirá um baixo custo operacional, pela seleção de fêmeas que parem precoce e regularmente e mostram harmonia com o ambiente, em termos de tamanho na idade adulta e adaptabilidade, por exemplo. Deste modo essa característica pode ser relacionada a precocidade à reprodução.

O mais importante no momento da escolha dos reprodutores é definir adequadamente os objetivos na produção dos animais, através do correto diagnóstico de seu rebanho e da definição de quais características devem ser melhoradas. Vale salientar que um reprodutor deve ter seu patrimônio genético o mais balanceado possível para atender às necessidades do criador. Alguns pais apresentam valores genéticos extraordinários para peso à desmama, por exemplo, mas um ganho de peso ao ano médio abaixo da média. O uso destes animais dará muita ênfase para determinada característica, em detrimento de outras. Cada criador, levando em consideração seu meio ambiente, seu mercado, as condições de manejo e pastagens, deverá definir seu limite ótimo para cada característica. Nem sempre o máximo é o melhor. Por exemplo, aumentar a produção leiteira ou o tamanho adulto das mães levará a uma maior necessidade de alimentos no futuro, para que os índices reprodutivos sejam mantidos. Neste sentido na Tabela 7 tem-se os melhores animais dentre os top 10 para as características quantitativa ( $Q$ ) e binária ( $B$ ) analisadas.

Tabela 7: Melhores animais dentre os top 10 para as características quantitativa ( $Q$ ) e binária ( $B$ ) analisadas.

Animal (código)	Característica ( $Q$ )	Característica ( $B$ )
2	1°	1°
17	4°	4°
280	6°	2°
188	7°	7°

Nota-se na Tabela 7 que os mesmos animais não são os melhores em cada característica, mas são animais elegíveis pois estão bem classificados para as características quantitativa ( $Q$ ) e binária ( $B$ ), devendo ser selecionados dependendo dos objetivos de seleção do criador.

## 6 Conclusão

O método de estimação da Máxima Verossimilhança Restrita apresentado através de modelos mistos para a característica quantitativa e modelos lineares generalizados mistos com distribuição binomial e utilizando a função de ligação *probit* para a característica binária mostrou-se adequado para estimar os componentes de variância e consequentemente estimar parâmetros genéticos das respectivas características.

As estimativas de herdabilidade apresentaram magnitudes variáveis, sendo alta para a característica quantitativa e moderada para a característica binária, evidenciando a existência de variabilidade genética aditiva entre os animais e que a seleção pode alcançar resultados consideráveis para o melhoramento de características de interesse econômico, promovendo ganhos genéticos significativos.

Os valores genéticos encontrados de todos os animais avaliados para as características quantitativa e binária, poderiam ser usados pelos criadores com o objetivo de atender à demanda do seu mercador consumidor, criando condições para o aumento do lucro e sustentação do rebanho através de ganhos genéticos, ao selecionar os animais do top 10.

## 7 Referências

- BARBIN, D. **Componentes de Variância: teoria e aplicações**. Piracicaba: FEALQ, 1993. 120p.
- CALUS, M. P., MULDER, H. A., VEERKAMP, R. F. **Estimating genomic breeding values and detecting QTL using univariate and bivariate models**. BMC Proceedings 2011, 5(Suppl 3): S5.
- CAMARINHA FILHO, J. A. **Modelos lineares mistos: estruturas de matrizes de variâncias e covariâncias e seleção de modelos**. 85 p. Tese (Doutorado em Agronomia), Universidade de São Paulo, Piracicaba 2002.
- CARNEIRO JÚNIOR, J. M., EUCLYDES, R. F., LOPES, P. S., TORRES, R. A. Avaliação de Métodos de Estimacão de Componentes de Variância Utilizando dados Simulados. **R. Bras. Zootec.**, v. 33, n. 2, p. 328-336, 2004.
- CARNEIRO JÚNIOR, J. M., ASSIS, G. M. L., EUCLYDES, R. F., MARTINS, W. M. O., WOLTER, P. F. Predicão de valores genéticos utilizando inferncia bayesiana e frequentista em dados simulados. **Acta Scientiarum. Animal Sciences.**, v. 32, n.32, p. 337-344, 2010.
- CORDEIRO, G.M., DEMÉTRIO, C.G.B. **Modelos Lineares Generalizados e Extensões**. Recife: UFRPE; Piracicaba: ESALQ, 2009. 254p.
- COSTA, R. G., ALMEIDA, C. C., PIMENTA FILHO, E. C., HOLANDA JUNIOR, E. V., SANTOS, N. M. Caracterizaão do sistema de produão caprino e ovino na regio semi-rida do estado da Paraba Brasil. **Arch. Zootec.** 57 (218): p. 195-205, 2008.
- CRUZ, C.D. **Princpios de gentica quantitativa**. Viosa: UFV, 2012. 394p.
- DE PAULA, M. C. **Heterogeneidade de varincias e interao gentipo x ambiente para produão de leite em bovinos da raa holandesa no estado do Paran**. 84 p. Tese (Doutorado em Zootecnia), Universidade Estadual de Maring, Maring, 2006.

- FALCONER, D.S. **Introdução à genética quantitativa**. Viçosa: UFV, 1987. 279p.
- FARIA, C. U., MAGNABOSCO, C. U., ALBUQUERQUE, L. G., REYES, A. L., BEZERRA, L. A. F., LOBO, R. B. Abordagem bayesiana e frequentista em análise genética tricaráter para crescimento e reprodução de bovinos nelore. **Ciência Animal Brasileira**, v. 9, n. 3, p. 598-607, 2008.
- FERREIRA, T. A. **Características morfológicas e de tipo, divergência e avaliação genética de caprinos leiteiros registrados no Brasil de 1976 a 2009**. 71 p. Dissertação (Mestrado em Zootecnia), Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina 2011.
- GARCIA, D.A., PEREIRA, I.G., SILVA, F.F., PIRES, A.V., PESSOA, M.C., OLIVEIRA, R.A.M. Modelos lineares generalizados mistos na avaliação genética da prenhez precoce na raça Nelore. **Arq. Bras. Med. Vet. Zootec.**, v.64, n.3, p. 675-682, 2012.
- HENDERSON, CR (1950) Estimation of genetic parameters. **Annals of Mathematical Statistics** 21: 309.
- KENNEDY, B. W. Variance component estimation and prediction of breeding values. **Canadian Journal of Genetics and Cytology**, v.23, n.4, p.565-578, 1981.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **J. R. Stat. Soc.**, Series A, v.135, p.370-384, 1972.
- PEREIRA, J.C.C. Melhoramento genético aplicado à produção animal. **Editora FEPMVZ**, Belo Horizonte, 5a Ed., p. 62-95, 2008.
- PATTERSON, H. D., THOMPSON, R. Recovery of interblock information when block sizes are unequal. **Biometrika**, Vol. 58, p. 545-554, 1971.
- R Development Core Team. **R: A language and environment for statistical computing**. Vienna: **R Foundation for Statistical Computing**, 2006. Disponível em: <http://www.R-project.org>. Acesso em: 11 mar.2013.
- RESENDE, M. D. V., BIELE, J. Estimação e predição em modelos lineares generalizados mistos com variáveis binomiais. **Rev. Mat. Estat.** , São Paulo, 20: p. 39-65, 2002.
- ROGES, D.L., IDALINO, R.C.L., SANTORO, K.R. Identificação dos polimorfismos e das similaridades existentes no segundo exon dos alelos do gene Bola-DRB3 da raça Bos Tauros. **19º Simpósio Nacional de Probabilidade e Estatística**. São Pedro, São Paulo. Resumo estendido. 2010.

ROSA, G. J. M. Métodos Quantitativos Aplicados ao Melhoramento Genético em Agricultura: **58<sup>a</sup> RBRAS e 15<sup>o</sup> SEAGRO**, 2013. 47 p. minicurso, Garden Hotel em Campina Grande – PB.

ROSANOVA, C. **Fatores favoráveis e limitantes ao desenvolvimento da cadeia produtiva da ovinocaprinocultura de corte no Brasil**. 42 p. Monografia (Pós-Graduação em Gestão Agro Industrial), Universidade Federal de Lavras, Minas Gerais, 2004.

SIMONATO, K. L. B. **Métodos da Máxima Verossimilhança e Máxima Verossimilhança Restrita na estimação de parâmetros do modelo de variabilidade espacial**. 104 p. Dissertação (Mestrado em Engenharia Agrícola), Universidade Estadual do Oeste do Paraná, Cascavel, 2009.

SZYDLOWSKI and PACZYNSKA. QTLMAS 2010: simulated dataset. **BMC Proceedings** 2011, 5(Suppl 3):S3

TORRES FILHO, R.A., TORRES, R.A., LOPES, P.S., PEREIRA, C. S., EUCLYDES, R. F., ARAÚJO, C. V., SILVA, M. A., BREDA, F. C. Estimativas de parâmetros genéticos para características reprodutivas de suínos. **Arq. Bras. Med. Vet. Zootec.**, v. 57, n. 5, p. 684-689, 2005.

TURKMAN, M. A. A., SILVA, G. L. **Modelos Lineares Generalizados – da teoria à prática** – Lisboa: Universidade de Lisboa, 2000. 151p.

VAYEGO, S. A. **Uso de modelos mistos na avaliação genética de linhagens de matrizes de frango de corte**. 104 p. Tese (Doutorado em Genética), Universidade Federal do Paraná, Curitiba, 2007.

VAZQUEZ, A.I., BATES, D.M., ROSA, G.J.M. et al. Technical note: an R package for fitting generalized linear mixed models in animal breeding. **J. Anim. Sci.**, v.88, p.497-504, 2010.

YAMAKI, M. **Estimação de parâmetros genéticos de produção de leite e de gordura da raça pardo-suíça, utilizando metodologias frequentista e bayesiana**. 64 p. Dissertação (Mestrado em Zootecnia), Universidade Federal de Viçosa, Minas Gerais, 2006.



## APÊNDICE A – Programa em linguagem R utilizado para a análise

```
setwd("")
### Indica o diretório para ler a página####

#Carregando os objetos

dados1=read.table("phenotypes.txt",h=T)
a=head(dados1)
dados2=read.table("pedigree.txt",h=T)
b=head(dados2)

##Comando para utilizar a tabela direto no latex

require(xtable)
xtable(a)
xtable(b)

#Carregando o pacote pedigreemm para análise dos dados.
require(pedigreemm)

#indicando quem é o animal, pai e mae no arquivo chamado pedigree.
ped = pedigree(sire=dados2$sire,dam=dados2$dam,label=as.character(dados2$animal))
head(ped)

# Realizando a análise da característica 1
ajuste1= pedigreemm(Trait1 ~ (1|animal) ,
                    data = dados1, pedigree = list(animal = ped))
```

```
#atenção: estrutura do pedigree!

#Resumo da característica 1
summary(ajuste1)

#Realizando a análise da característica 2
ajuste2= pedigreemm(Trait2 ~ 1+(1|animal) ,
                    data = dados1, pedigree = list(animal = ped),
                    family = 'binomial' (link = 'probit'))

#Resumo da característica 2
summary(ajuste2)

#Utilizando o comando para extrair os efeitos aleatórios

i=raneef(ajuste1)$animal
j=raneef(ajuste2)$animal

#Estatística descritiva dos efeitos

summary(i)
summary(j)

#Extraindo a quantidade de efeitos que eu desejar

indi=which(i> 12)
indi
edit(i)

ind = which(j>0.90)
ind
edit(j)

k=as.vector(i)
is.vector(k)
```

```
head(k)
```

```
order(i)
```

```
#atenção: estrutura do pedigree!
```

```
summary(ajuste2)
```