

UNIVERSIDADE FEDERAL DE MINAS GERAIS

NATASCHA ALMEIDA MARQUES DA SILVA

**SELEÇÃO DE MODELOS DE REGRESSÃO
NÃO LINEARES E APLICAÇÃO DO ALGORITMO
SAEM NA AVALIAÇÃO GENÉTICA DO
CRESCIMENTO DE BOVINOS NELORE**

BELO HORIZONTE
MINAS GERAIS
2010

NATASCHA ALMEIDA MARQUES DA SILVA

**SELEÇÃO DE MODELOS DE REGRESSÃO NÃO LINEARES E
APLICAÇÃO DO ALGORITMO SAEM NA AVALIAÇÃO GENÉTICA
DO CRESCIMENTO DE BOVINOS NELORE**

Tese apresentada ao Programa de Pós- Graduação em Zootecnia da Escola de Veterinária da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Doutor em Zootecnia, área de concentração Genética e Melhoramento animal.

Orientadora: Profa. Dra. Ângela Maria Quintão Lana (UFMG)

Co-orientadores: Prof. Dr. Fabyano Fonseca e Silva (UFV)

Prof. Dr. José Aurélio Garcia Bergmann (UFMG)

Tese defendida e aprovada em 09/07/2010, pela Comissão Examinadora constituída por:

Profa Dra Ângela Maria Quintão Lana
(Orientadora)

Prof. Dr. Martinho de Almeida e Silva

Prof. Dr. Fábio Luiz Buranelo Toral

Dr. Miguel Houri Neto

Prof. Dr. Renato Ribeiro de Lima

DEDICATÓRIA

Dedico este trabalho ao Prof. Luiz Henrique de Aquino (Caveira), que me ensinou a amar a pesquisa científica e a estatística. Me ensinou o que é ser ético e profissional. E a cima de tudo, o que é ser um educador!

AGRADECIMENTOS

À meus pais Francisco e Eduarda e meus irmãos Thais e Pedro: pelo amor, apoio e paciência em todos os momentos da minha vida. Vocês são essenciais para mim!!!! Amo muito vocês!!!!

Aos meus tios, tias, primos, primas e avó Gerusa, pela presença constante e torcida fiel. A vida com vocês por perto é sempre melhor!!!!

À minha amiga, professora e orientadora Ângela, pelos ensinamentos, companheirismo, carinho e verdadeira amizade em todas as horas. Obrigada por ter sempre acreditado e confiado em mim!

Ao meu amigo, pai e irmão Miguel por ser o anjo bom que Deus mandou para minha vida. Seus conselhos, suas palavras, seu carinho e sua amizade foram fundamentais para eu chegar até aqui. Obrigada!!!

Ao grande amigo e professor Fabyano. Por ter estado presente por tantos anos de forma tão constante e alegre. Pelos grandes ensinamentos e pela sempre disposição em ajudar. Obrigada sempre!!!!

Aos Doutores Martinho de Almeida e Silva, Fábio Luiz Buranelo Toral, Renato Ribeiro de Lima e Miguel Houri Neto, por aceitarem participar da banca de defesa. Obrigada pelas ótimas sugestões.

Ao professor José Aurélio pelos grandes ensinamentos e pela força de sempre!!!!

As amigas-irmãs, Carol, Cris, Lídia, Lú e Lilian. Pelas maravilhosas quartas feiras... Pela amizade verdadeira e eterna!

Aos distantes e sempre presentes amigos de Lavras. Vocês fazem parte desta conquista.

Aos amigos da UFMG. Pelos bons momentos vividos. Pelas lindas amizades construídas.

A minha primeira, grande e eterna amiga Polly. Sempre juntas! Te amo!

Ao CNPq pelos recursos financeiros concedidos.

A todos que contribuíram para essa vitória, muito obrigada!!!!

SUMÁRIO

CAPÍTULO 1: SELEÇÃO E CLASSIFICAÇÃO MULTIVARIADA DE MODELOS DE CRESCIMENTO NÃO LINEARES PARA BOVINOS NELORE

RESUMO	8
ABSTRACT	8
1. INTRODUÇÃO	9
2. REVISÃO DE LITERATURA	9
2.1 MODELOS MATEMÁTICOS	9
2.1.2 Modelos lineares, linearizáveis ou não lineares.....	10
2.1.3 Modelos estáticos ou dinâmicos.....	10
2.1.4 Modelos determinísticos ou estocásticos.....	11
2.1.5 Modelos empíricos ou mecanísticos	11
2.2 MODELOS NÃO LINEARES DE CRESCIMENTO	12
2.3 AVALIADORES DE QUALIDADE DE AJUSTE	14
2.4 ANÁLISE DE AGRUPAMENTO	16
2.4.1 Métodos não hierárquicos.....	16
2.4.2 Métodos hierárquicos.....	16
2.5 DETERMINAÇÃO DO NÚMERO ÓTIMO DE GRUPOS	17
3. MATERIAL E MÉTODOS	18
3.1 DADOS	18
3.2 ESTIMAÇÃO DOS PARÂMETROS DOS MODELOS NÃO LINEARES	18
3.2.1 Método do mínimos quadrados ordinários.....	18
3.2.2 Método de Gauss Newton.....	19
3.3 AVALIADORES DE QUALIDADE DE AJUSTE	20
3.4 ANÁLISE DE AGRUPAMENTO	22
3.5 DETERMINAÇÃO DO NUMERO ÓTIMO DE GRUPOS E ESCOLHA DO MELHOR MODELO	23
3.6 ASPECTOS COMPUTACIONAIS	24
4. RESULTADOS E DISCUSSÃO	24
5. CONCLUSÕES	28
6. REFERÊNCIAS BIBLIOGRÁFICAS	29

CAPÍTULO 2: PROPOSTA DE ANÁLISE GENÉTICA DE CURVAS DE CRESCIMENTO DE BOVINOS POR MEIO DO ALGORITMO SAEM

RESUMO	34
ABSTRACT	34
1 INTRODUÇÃO	35
2 REVISÃO DE LITERATURA	36
2.1 SELEÇÃO DE CURVAS EM BOVINOS	36
2.2 MODELOS DE REGRESSÃO NÃO LINEAR	36
2.3 MODELOS MISTOS	37
2.4 ESTIMAÇÃO DOS COMPONENTES DE VARIÂNCIA	38
2.5 ALGORITMO EM (ESPERANÇA – MAXIMIZAÇÃO)	39
2.6 ALGORITMO SAEM	40
3 MATERIAL E MÉTODOS	41
3.1 DADOS	41

3.2	MODELO DE BRODY MODIFICADO.....	42
3.3	MODELO TOURO.....	42
3.4	OBTENÇÃO DOS COMPONENTES DE VARIÂNCIA	42
3.4.1	Algoritmo SAEM	43
3.4.2	Two Step.....	44
3.5	VALORES GENÉTICOS E CURVAS GENÉTICAS DE TOUROS	44
3.5.1	Metodologia adaptada de Pong-Wong e Hadjipavlou (2009).....	45
3.5.2	Análises unicaracterísticas para pesos observados em idades padrão	46
3.6	ANÁLISES REALIZADAS	46
3.7	ASPECTOS COMPUTACIONAIS	47
4	RESULTADOS E DISCUSSÃO.....	47
4.1	COMPONENTES DE VARIÂNCIA E PARÂMETROS GENÉTICOS	47
4.2	VALORES GENÉTICOS PARA OS PARAMETROS DO MODELO	49
4.3	CURVAS GENÉTICAS E VALORES GENÉTICOS EM IDADES PADRÃO	53
5	CONCLUSÕES.....	57
6	IMPLICAÇÕES	57
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	57

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1 - Modelos de regressão não-linear para descrever curvas de crescimento.....	13
Tabela 2 - Estimativas dos parâmetros para os modelos de crescimento de bovinos Nelore considerando o ajuste individual.....	25
Tabela 3 - Resultados dos avaliadores da qualidade de ajuste para os modelos não lineares considerando ajuste individual.....	26
Tabela 4 - <i>Clusters</i> com os seus respectivos modelos e médias dos avaliadores.....	28

CAPÍTULO 2

Tabela 1- Matrizes de co(variância) dos efeitos genéticos e residuais e médias estimadas dos parâmetros A e K para os quatro análises realizadas.....	47
Tabela 2- Valores calculados de herdabilidades (h^2_A e h^2_K) e correlação genética ($r_g(A, K)$) para os parâmetros A e K	49
Tabela 3 – Ordenação dos dez touros superiores e inferiores para os parâmetros A e K nas 4 análises realizadas.....	52
Tabela 4- Concordância dos métodos de análise considerando comparações 2 a 2 na ordenação dos animais.....	52

Tabela 5- Valores do coeficiente de correlação entre valores genéticos das idades padrão, nas quatro análises realizadas.....	56
Tabela 6- Valores do coeficiente de correlação de Spearman entre os valores genéticos das idades padrão dos touros pela análise unicaracterística e pelas metodologias testadas.....	56

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1 - Ilustração da determinação geométrica do ponto de máxima curvatura da curva representativa do comportamento do RMSSTD em função do número de grupos.....	24
Figura 2 - Ajuste exponencial do índice RMSSTD em função do número de cluster.....	27
Figura 3 – Dendrograma resultante da análise de agrupamento.....	27

CAPÍTULO 2

Figura 1- Gráficos de dispersão e valores do coeficiente de Spearman entre os valores genéticos preditos para o parâmetro A (Ua) nas quatro análises realizadas.....	50
Figura 2- Gráficos de dispersão e valores do coeficiente de Spearman entre os valores genéticos preditos para o parâmetro K (Uk) nas quatro análises realizadas.....	51
Figura 3 - Curvas genéticas individuais para touros e curva média obtidas pela análise SAEM1.....	53
Figura 4 - Curvas genéticas individuais para touros e curva média obtidas pela análise TS1.....	54
Figura 5 - Curvas genéticas individuais para touros e curva média obtidas pela análise SAEM2.....	54
Figura 6- Curvas genéticas individuais para touros e curva média obtidas pela análise TS2.....	55

CAPÍTULO 1

SELEÇÃO E CLASSIFICAÇÃO MULTIVARIADA DE MODELOS DE CRESCIMENTO NÃO LINEARES PARA BOVINOS NELORE

Resumo

O objetivo principal desse trabalho foi utilizar a análise de agrupamento para classificar e selecionar modelos não lineares de crescimento de bovinos Nelore tendo em vista os resultados de diferentes avaliadores de qualidade de ajuste. Ajustaram-se 12 modelos não-lineares, cuja qualidade de ajuste foi medida pelo coeficiente de determinação (R^2), quadrado médio do erro (QME), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC), erro quadrático médio de predição (MEP) e coeficiente de determinação de predição (R^2_p). O modelo Brody foi o que apresentou o melhor ajuste para o conjunto de dados.

Palavras-chave: Análise multivariada, modelos não lineares, bovinos Nelore

SELECTION AND MULTIVARIATE CLASSIFICATION OF NONLINEAR GROWTH MODELS FOR NELORE CATTLE.

Abstract

This study aimed to evaluate cluster analysis in classifying and selecting non linear models to describe Nelore beef cattle growth based on different goodness of fit criteria tests. A total of 12 non linear models were evaluated based on the following criteria: determination coefficient (R^2), error mean square (QME), Akaike information criterion (AIC), Bayesian information criterion (BIC), mean quadratic error of prediction (MEP) and predicted determination coefficient (R^2_p). The Brody model showed the best goodness of fit for this data set.

Key-Words: multivariate analysis, nonlinear models, Nelore cattle

1 - Introdução

Atualmente, uma forma prática e consistente de analisar a eficiência produtiva de raças de bovinos de corte é por meio do estudo de curvas de crescimento, pois estas representam uma trajetória longitudinal dos pesos apresentados pelos animais em função do tempo.

Modelos de regressão não-linear, desenvolvidos empiricamente para relacionar dados peso-idade, têm-se mostrado adequados para descrever curvas de crescimento. Esses modelos apresentam parâmetros com interpretação biológica e assim possibilitam ter um maior entendimento a respeito do processo de crescimento dos animais. Esses parâmetros são caracterizados principalmente pelo peso à maturidade, que representa o peso na idade adulta, e pela velocidade de crescimento, que representa uma medida de precocidade.

Existem alguns requisitos que devem ser seguidos para que um modelo matemático não-linear seja descritivo de uma relação peso-idade. Estes requisitos se baseiam nos seguintes itens: interpretação biológica dos parâmetros (adequação a realidade), ajuste com pequenos desvios (precisão) e grau de dificuldade do ajuste (facilidade computacional).

A literatura apresenta vários modelos que seguem esses requisitos e descrevem o crescimento dos bovinos, sendo que cada um desses modelos apresenta suas vantagens e desvantagens sob pontos de vista estatístico. Dessa forma, pesquisadores e produtores podem se questionar qual desses modelos é o mais apropriado para descrever o crescimento corporal de uma população de bovinos. A resposta para tal questionamento pode ser dada mediante informações provenientes de avaliadores da qualidade de ajuste, os quais nos permite indicar estatisticamente o melhor modelo. Porém, na presença de diferentes populações, as quais podem ser denominadas de tratamentos sob o ponto de vista experimental, muitas vezes o modelo que melhor representou o crescimento dos animais de uma população pode não ser o melhor para as demais, ou seja, os resultados dos avaliadores podem variar de acordo com a população considerada. Na área de genética e melhoramento animal estas populações, geralmente são designadas por diferentes raças, ou ainda diferentes combinações de raças, sendo estas denominadas de grupos genéticos.

Assim, uma análise que considere simultaneamente todos esses avaliadores de qualidade seria necessária para estabelecer qual modelo melhor se ajusta aos dados de crescimento. Uma proposta interessante seria a utilização de métodos de classificação multivariados, como a análise de agrupamento, ou análise de cluster, a qual permite a formação de grupos homogêneos de acordo com a avaliação simultânea de várias características de interesse. A utilidade desta metodologia reside na possibilidade de classificar, ou agrupar, modelos cujos resultados dos avaliadores são concordantes.

Assim, o objetivo do trabalho foi utilizar a análise de agrupamento para classificar os modelos de crescimento não-lineares e indicar qual seria o de melhor ajuste tendo em vista os resultados de diferentes avaliadores de qualidade de ajuste, considerando dados de bovinos Nelore.

2-Revisão de Literatura

2.1 – Modelos Matemáticos

De acordo com Rondon et al. (2002), um “modelo” é a descrição matemática por meio de equações, de fenômenos biológicos como: crescimento, digestão, absorção de nutrientes, entre outros. Assim, variáveis quantitativas são tomadas para representar os fatores que influenciam o fenômeno em estudo.

O sucesso de um modelo dependerá da seleção adequada dos fatores causativos e da percepção de como cada um desses fatores atua sobre a resposta alvo. Os elementos numéricos de um modelo matemático são ditos parâmetros. Geralmente eles exprimem a influência de cada fator sobre a resposta medida (Sampaio, 2002).

Mendes (2007) afirma que uma atividade muito comum em análises estatísticas é a avaliação da relação entre uma variável dependente e uma ou mais variáveis independentes. Essa análise pode ser feita por meio dos modelos de regressão, os quais se dividem em duas classes distintas: os lineares e os não lineares. Segundo Sampaio (2002), o termo linearidade se refere aos parâmetros existentes no modelo, e não ao aspecto gráfico do mesmo.

2.1.1 - Modelos lineares, linearizáveis ou não-lineares

De acordo com Mendes (2007), os modelos podem ser classificados como: modelos lineares, linearizáveis e não lineares.

- a) Os modelos lineares são aqueles que são lineares em relação aos parâmetros, isto é, as derivadas parciais não dependem dos parâmetros, logo:

$$\frac{\partial}{\partial \theta_j} f_i(X, \theta) = g(X)$$

com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$; em que n é o número de observações e p é o número de parâmetros do modelo.

- b) Os modelos linearizáveis são aqueles que podem ser transformados em lineares por meio de alguma transformação. Para o modelo:

$$y = a^X \cdot e$$

no qual o erro é multiplicativo, aplicando-se o logaritmo a igualdade, tem-se:

$$\begin{aligned} \ln(Y) &= \ln(a^X \cdot e) \\ \ln(Y) &= \ln(a^X) + \ln(e) \\ \ln(Y) &= X \ln(a) + \ln(e) \end{aligned}$$

Sendo $Z = \ln(y)$; $b = \ln(a)$; $e^* = \ln(e)$, tem-se:

$$Z = Xb + e^*, \text{ que é linear, pois:}$$

$$\frac{\partial Z}{\partial b} = X = g(X). \text{ Logo o modelo é dito linearizável.}$$

- c) Os modelos não lineares são aqueles que não se enquadram nos casos anteriores. Para o modelo:

$$y = a^X + e$$

o erro é aditivo e não existe nenhuma transformação capaz de o tornar linear. Assim a derivada do modelo em relação ao parâmetro a depende do parâmetro, isto é,

$$\frac{\partial f}{\partial a} = Xa^{X-1} = g(X, a). \text{ Portanto o modelo é dito não linear.}$$

Segundo Baldwin (1995), os modelos matemáticos ainda podem ser classificados como: estáticos ou dinâmicos, determinísticos ou estocásticos e como empíricos ou mecanísticos.

2.1.2 - Modelos estáticos ou dinâmicos

Um modelo pode ser chamado de estático se os dados são obtidos somente em um certo momento do tempo, ou seja, os modelos estáticos descrevem o fenômeno em determinado momento ou instante. Um modelo é dito dinâmico quando as respostas são mensuradas sequencialmente no tempo, ou seja, as respostas podem variar no tempo. As curvas de crescimento são exemplos de modelos dinâmicos.

De acordo com Baldwin (1995), os modelos dinâmicos, são baseados em equações diferenciais, que são integradas ao longo do tempo, normalmente usando uma técnica numérica. Esses modelos são capazes de identificar os impactos dos tempos anteriores e atuais sobre os desempenhos subsequentes. Já os modelos estáticos estão normalmente na forma de equações algébricas que são solucionadas por um conjunto específico de condições em determinado tempo.

2.1.3 - Modelos determinísticos ou estocásticos

Modelos determinísticos são aqueles que possuem uma única resposta. O termo “determinístico” implica que todas as soluções de uma equação ou de um conjunto de equações são exatas. Em ciência animal, o uso deste termo implica que as soluções são especificamente aplicadas para a resposta média dos animais de uma população (Baldwin, 1995).

Os modelos estocásticos podem também ser chamados de probabilísticos, pois, as possíveis respostas dependem de uma distribuição de probabilidades. Estes modelos permitem estimar, além das respostas dos animais, os parâmetros de dispersão, os quais medem a variabilidade dos indivíduos que da população. A distribuição de probabilidades mais comumente usada neste tipo de modelo é a distribuição normal (Rondon et al., 2002).

Na área de nutrição e manejo, a maioria dos modelos disponíveis atualmente é determinístico, enquanto que os modelos estocásticos são os mais usados nos estudos de melhoramento animal (Rondon et al., 2002). A representação estocástica é utilizada, quando um limitado número de observações é avaliado, pois dessa forma é mais difícil detectar o real comportamento de uma população. No entanto, quando um grande número de observações é avaliado num experimento, a variação amostral dos valores médios é menor e um modelo determinístico obterá uma boa aproximação da realidade (Sampaio, 1988).

2.1.4 - Modelos empíricos ou mecanísticos

Os modelos empíricos são baseados apenas em correlações ou associações entre duas ou mais variáveis, sem levar em consideração os mecanismos que controlam o fenômeno. Por outro lado, os modelos mecanísticos tentam explicar ou descrever os mecanismos envolvidos nos fenômenos, baseando-se nas leis da física, química, bioquímica etc. (Rondon et al., 2002).

Segundo Ferreira Filho (2009), os modelos empíricos buscam apenas descrever os fenômenos em estudo, enquanto os mecanísticos tentam explicar estes fenômenos. Segundo este autor, apenas conseguir descrever pode ser em muitos casos uma tarefa ambiciosa demais. Assim, a modelagem empírica já se dá por satisfeita se é capaz de descrever o processo estudado na região experimental investigada. Isto quer dizer que modelos empíricos são também modelos locais.

A associação do peso de um animal com a idade em cada tempo é uma situação clássica de modelo empírico. No entanto, imaginar que o peso do animal, apresenta necessidade de manutenção, consumo e conversão alimentar em determinada idade é mais coerente, sendo exemplo de possível modelo mecanístico (Sampaio, 2002),

Rondon et al. (2002) afirmam que a dificuldade em se definir os mecanismos envolvidos nos fenômenos biológicos faz com que a maioria dos modelos propostos sejam empíricos, pelo fato deste tipo de modelagem não exigir o conhecimento sobre os processos intermediários que produzem a resposta final e assim conseguir a simplificação do modelo usado para o estudo.

Os modelos geralmente só descrevem o crescimento do ponto de vista da composição corporal. Entretanto, é muito importante tentar desenvolver modelos mecanísticos que descrevam fenômenos como o metabolismo de nutrientes, a digestão, o crescimento de tecidos e o desenvolvimento embrionário, para se aprofundar nos conhecimentos dos processos envolvidos.

Segundo Sampaio (1988), um modelo pode ser definido por meio da combinação de qualquer uma das classificações acima, embora somente algumas sejam possíveis de ocorrer. No contexto da classificação dos modelos discutidos, as equações de regressão, em geral, são consideradas estáticas,

estocásticas e empíricas. No entanto, situações que exigem todos os valores especificados sejam exatos, serão consideradas como empíricas, estáticas e determinísticas. Já a situações em que o modelo é composto de equações diferenciais que capturam o conhecimento das funções subjacentes serão mecanísticos, determinísticos e dinâmicos (Baldwin, 1995).

2.2 – Modelos não lineares de crescimento

Segundo Forni (2008), o desenvolvimento de teorias e técnicas para o ajuste de modelos matemáticos que descreve o crescimento ocorreu de forma multidisciplinar, uma vez que o conhecimento dos aspectos gerais deste processo desperta interesse em várias áreas de conhecimento.

SANDLAND (1983) relatou que uma das primeiras tentativas de descrever o crescimento por meio de funções matemáticas foi realizada em 1835 por Quetelet que estabeleceu analogias entre diferentes fases do crescimento de um organismo propondo uma curva sigmóide para descrevê-lo.

De acordo com Fitzhugh (1976), uma curva sigmóide descrevendo uma sequência de medidas de tamanho, frequentemente peso corporal, sugere o termo “curva de crescimento” ou “curvas de peso-idade”. O autor destaca que os principais objetivos no ajuste de curvas de crescimento são descritivos, de modo que a informação contida numa sequência de pontos peso-idade é reduzida em relativamente poucos parâmetros que tenham interpretação biológica.

Segundo Sampaio (1988), as tentativas iniciais de modelar o crescimento animal foi a partir de modelos empíricos, determinísticos e dinâmicos, sendo o peso vivo o principal componente desta modelagem estando este relacionado com a idade. Segundo este autor, os modelos estocásticos foram introduzidos por Sandland e McGilchrist em seu trabalho em 1979. Estes autores se basearam em uma extensão do modelo logístico com a adição de um elemento estocástico. De acordo com Rondon et al. (2002), o estudo da curva de crescimento de cada animal, por meio de modelos estocásticos, tem se mostrado bastante útil, principalmente nas pesquisas de melhoramento animal. Com esses modelos é possível ter uma melhor avaliação da população, podendo-se planejar maneiras de se promover mudanças desejáveis na forma da curva de crescimento dos animais, o que poderia ser feito por meio de manejos de nutrição e/ou seleção de animais (Gous, 1998).

Os parâmetros da curva de crescimento são utilizados para prever taxas de crescimento, necessidades alimentares, peso à maturidade, graus de maturidade, e servindo como critérios de seleção para programas de melhoramento animal. Nesse sentido, as curvas de crescimento refletem a relação entre a idade do animal e o seu impulso de crescimento e maturidade, sendo importantes para pesquisas e recomendações sobre eficiência de produção em programas de melhoramento, contribuindo, dessa maneira, para aumentar o lucro do produtor (Mendes, 2006).

Uma das formas de descrever as curvas de crescimento é por meio de modelos de regressão não linear que podem ser representados da seguinte forma:

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, i = 1, \dots, n$$

em que:

y_i representa a observação da variável dependente designada variável resposta,

x_i representa a variável independente designada variável explicativa,

$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]'$ é um vetor de parâmetros p dimensional desconhecido

$f(x_i, \boldsymbol{\theta})$ é uma função de variáveis regressoras e de parâmetros conhecida como função esperança ou função de regressão,

ε_i representa o erro aleatório não observável suposto independente e identicamente distribuído, normal com média zero e variância desconhecida σ^2 .

Existem na literatura vários modelos de regressão não-linear que podem ser utilizados para descrever curvas de crescimento animal. Mas segundo vários autores, entre eles, Fitzhugh Jr (1976), Oliveira et al. (2000), Mazzini et al. (2003), Freitas (2005), Mendes (2007) e Silva et al. (2010), os

modelos da família Richards (Brody, Von Bertalanffy, Richards, Logística e Gompertz), são os mais utilizados para descrever o crescimento de bovinos especificamente.

Tabela 1- Modelos de regressão não-linear para descrever curvas de crescimento

Modelo de Curva de Crescimento	Referência
$y_i = \frac{\beta_1}{\left(1 + \beta_4 e^{(\beta_3 \beta_2 - x_i)}\right)^{\frac{1}{\beta_4}}} + e_i$	Schnute (1981)
$y_i = \beta_1 \left(1 - e^{(\beta_3 \beta_2 - \beta_3 x_i)}\right) + e_i$	Mitscherlich (1919)
$y_i = \frac{\beta_1}{\left(1 + e^{(\beta_2 - \beta_3 x_i)}\right)^{\frac{1}{\beta_4}}} + e_i$	Richards (1959)
$y_i = \beta_1 e^{(-e^{(\beta_2 - \beta_3 x_i)})} + e_i$	Gompertz (GOMPERTZ e PHILOS, 1825)
$y_i = \frac{\beta_1}{\left(1 + e^{(\beta_2 - \beta_3 x_i)}\right)} + e_i$	Logístico (RATKOWSKI, 1983)
$y_i = \beta_1 - \beta_2 e^{(-\beta_3 x_i)} + e_i$	Meloun I (MELOUN e MILITKY, 1996)
$y_i = \beta_1 - e^{(-\beta_2 - \beta_3 x_i)} + e_i$	Meloun II (MELOUN e MILITKY, 1996)
$y_i = (\beta_1 - \beta_2 e^{-\beta_3 x_i}) + e_i$	Brody (1945)
$y_i = \beta_1 \left(1 - \beta_2 e^{-\beta_3 x_i}\right)^3 + e_i$	von Bertalanffy (1957)
$y_i = \frac{\beta_1 x_i}{x_i + \beta_2} + e_i$	Michaelis-Menten (MICHAELIS e MENTEN, 1913)
$y_i = \frac{\beta_2 \beta_3^{\beta_4} + \beta_1 x_i^{\beta_4}}{\beta_3^{\beta_4} + x_i^{\beta_4}} + e_i$	Michaelis-Menten Modificado (LOPEZ et al., 2000)
$y_i = \beta_1 - \beta_2 e^{-e^{\beta_3 x_i \beta_4}} + e_i$	Weibull (RATKOWSKI, 1983)

Na Tabela 1 estão apresentadas as funções $f(x_i, \theta) = f(x_i, \beta_1, \beta_2, \dots, \beta_p)$ de alguns modelos de regressão não-linear utilizados para descrever curvas de crescimento animal em geral. (Silveira, F. G, 2010). De forma geral, esses modelos têm por objetivo descrever uma trajetória assintótica da variável dependente peso, em função da variável independente tempo. Geralmente, a diferença entre tais modelos é dada pela definição do ponto de inflexão da curva, que confere uma forma sigmóide a mesma. Porém, para alguns modelos este ponto pode não existir

Para os modelos apresentados (tabela 1), o parâmetro β_1 representa o peso adulto, ou peso assintótico do animal. O parâmetro β_3 representa a taxa de maturidade, ou velocidade de crescimento, exceto para os dois últimos modelos, Michaelis-Menten Modificado e Weibull, em que para o primeiro este parâmetro representa a idade na qual se atinge a metade do peso adulto e no segundo a taxa de maturidade é obtida como sendo uma função dos demais parâmetros. Os modelos que apresentam o parâmetro β_4 possuem ponto de inflexão variável, cuja localização é determinada pelo parâmetro em questão. Os demais modelos ou apresentam o ponto de inflexão fixo, ou não o possuem, como é o caso dos modelos Brody e Michaelis-Menten. De forma geral, não há uma interpretação prática para o parâmetro β_2 , sendo este uma constante de integração.

Perotto et al. (1992) citado por Elias (1998), relata que Brody, Logístico, Gompertz e von Bertalanffy são casos especiais do modelo Richards. Este último por ter o parâmetro β_4 flexível, o qual determina o ponto de inflexão, pode assumir a forma dos outros quatro modelos em questão. Devido ao fato de apresentar um parâmetro a mais, o modelo Richards geralmente apresenta excelentes ajustes, porém é de convergência difícil.

O modelo Weibull também apresenta quatro parâmetros como o Richards, porém, segundo Rodrigues et al. (2007), ao ajustar estes dois modelos a dados de cordeiro da raça Bergamácia, concluiu por meio do cálculo de medidas de curvatura que o modelo Richards é estatisticamente muito mais complexo que o modelo Weibull, o que implica em uma baixa praticidade em razão à dificuldade de convergência.

Quanto a utilização do modelo Michaelis-Menten na descrição de curvas de crescimento, Machado et al. (2005) ajustou o mesmo a dados de novilhos Nelore por meio de três diferentes metodologias de ajuste, considerando a forma original não-linear e duas formas linearizadas. Foi observada uma superioridade do método de ajuste não-linear, pois este além de apresentar baixos valores para o Quadrado Médio do Erro combinado e facilidade de convergência, proporcionou também estimativas dos parâmetros mais condizentes com a realidade. Lopez et al. (2000) apresentaram uma generalização do modelo original de Michaelis-Menten com o objetivo de descrever curvas de crescimento sigmóides, uma vez que o modelo original não apresenta ponto de inflexão e concluíram que o modelo proposto ajustou-se muito bem a dados de peso-idade de várias espécies animais, apresentando resultados similares aos modelos Gompertz e Richards.

Os modelos Schnute, Meloun I e II e Mitscherlich, mostrados na Tabela 1, foram utilizados por Meloun e Militky (1996) para descrever o crescimento de uma estrutura auricular de fetos, e devido a grande flexibilidade apresentada, os mesmos podem se caracterizar como modelos alternativos para descrição de curvas de crescimento de animais.

2.3 – Avaliadores de qualidade de ajuste

Em razão das diferenças entre os modelos de regressão não-linear, quando estes são ajustados a um mesmo conjunto de dados, torna-se necessário utilizar metodologias estatísticas com o intuito de compará-los e indicar o melhor modelo (SILVEIRA et al., 2009). Tais técnicas são denominadas de avaliadores da qualidade de ajuste.

Na literatura, a maioria dos trabalhos envolvendo comparação de modelos de crescimento utiliza apenas o coeficiente de determinação (R^2) e o desvio padrão residual como avaliadores da qualidade de ajuste. Tal procedimento pode ser explicado pelo fato destes avaliadores serem apresentados diretamente nos arquivos de saídas de softwares estatísticos, como SAS[®] e R (SILVEIRA et al., 2009).

De acordo com Silveira et al. (2009), se outros importantes avaliadores como o critério de informação de Akaike (AIC), (AKAIKE, 1974), critério de informação bayesiano, BIC, (SCHWARZ, 1978), erro de predição médio (EPM), coeficiente de determinação predito e percentual de convergência também forem considerados, a indicação do melhor modelo pode ser mais precisa, uma vez que estes avaliadores levam em consideração outros fatores como a análise da independência residual e o grau de parametrização dos modelos comparados.

Em um estudo com bovinos Nelore, Silva et al (2004) encontram a função de Gompertz como a mais adequada para descrever o crescimento dos animais. Para a determinação da melhor função foram utilizados os seguintes avaliadores de qualidade: porcentagem de convergência, coeficiente de determinação ajustado, autocorrelação residual de primeira ordem e estatística de Durbin e Watson. Todas as funções apresentaram comportamentos semelhantes, mas a função de Gompertz foi escolhida a melhor por apresentar bom ajuste, maior porcentagem de convergência e estimativas com interpretação biológica.

Garnero et al. (2005) utilizaram o coeficiente de determinação, o número de iterações e o quadrado médio do erro como avaliadores de qualidade de ajuste e concluíram, em um estudo com fêmeas Nelore, que o modelo de Brody foi o mais indicado para estabelecer padrões médios de crescimento.

Mazzini et al. (2005), em um estudo de curvas de crescimento de novilhos Hereford, utilizaram o percentual de convergência, o quadrado médio do resíduo, o coeficiente de determinação ajustado e o erro de predição médio para comparar os modelos de crescimento utilizados. Esses autores chegaram à conclusão que o coeficiente de determinação ajustado não era um bom avaliador, pelo fato da diferença entre eles ser irrisória no estudo. Assim, concluíram que o modelo de Von Bertalanffy e o de Gompertz foram os melhores para descrever o crescimento dos animais baseados apenas no percentual de convergência, no quadrado médio do resíduo e no erro de predição médio. Esse resultado concorda com Oliveira et al. (2000), que também concluíram que o coeficiente de determinação ajustado apresenta diferença mínima entre os modelos. Neste estudo, os autores consideraram o modelo de Von Bertalanffy como o melhor no ajuste de curvas de crescimento de fêmeas da raça Guzerá.

Segundo Paz et al. (2004), o modelo Logístico foi o que melhor descreveu o crescimento animal de bovinos de corte em três diferentes grupos genéticos. Em seu estudo ele comparou os modelos de acordo com as dificuldades computacionais (% de convergência e número de iterações) e qualidade do ajuste (coeficiente de determinação e quadrado médio do resíduo).

Freitas (2005) ajustou os modelos de crescimento Brody, Gompertz, Richards, Logístico e Von Bertalanffy a dados de peso-idade de oito espécies animais e concluiu que o modelo Logístico, seguido do modelo de Von Bertalanffy, foram os mais versáteis para descrever os dados de crescimento das espécies animais estudadas. Nesse estudo foi utilizado como avaliador de qualidade o coeficiente de determinação, a porcentagem de convergência e a interpretabilidade biológica dos parâmetros. Esse autor conclui também que de modo geral os modelos ajustados tenderam a superestimar os pesos iniciais e finais em todas as espécies animais estudadas.

A comparação entre modelos quanto ao seu ajustamento aos dados pressupõe o teste da hipótese de que todos os modelos são igualmente apropriados para descrever a trajetória de crescimento. Embora testes estatísticos objetivos para não rejeitar ou rejeitar essa hipótese fossem desejáveis, a estrutura dos dados e a forma como são analisados não permitem a condução de um teste definitivo (OLIVEIRA et al., 2000).

Segundo Silveira et al. (2009), quanto maior o número de avaliadores considerados, mais adequada é a indicação dos melhores modelos. Porém, quando grande número de avaliadores é considerado, a escolha dos modelos se transforma em um processo complexo, uma vez que um mesmo modelo pode apresentar alta performance para um avaliador e baixa para outros. Uma proposta para minimizar esse problema, seria utilizar métodos multivariados de classificação, como análise de agrupamento, com o intuito de agrupar modelos, cujos resultados fossem semelhantes simultaneamente em relação a todos os avaliadores considerados. Assim seria possível escolher um

modelo que se encontra dentro do grupo definido como de alta performance, ou seja, escolher aquele modelo que apresentou melhor ajuste, para todos os avaliadores considerados simultaneamente.

2.4 – Análise de agrupamento

A análise de agrupamento constitui uma metodologia numérica multivariada que, segundo Mingoti (2007), tem como objetivo dividir os elementos da amostra (n), ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com as variáveis que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características.

Uma questão importante em análise de agrupamento refere-se ao critério a ser utilizado para se decidir até que ponto dois elementos do conjunto amostral podem ser considerados como semelhantes ou não. De acordo com Cruz et al. (2004), existem dois tipos de medidas que podem ser consideradas como critérios para a caracterização da semelhança entre os elementos amostrais. São as medidas de similaridade e as medidas de dissimilaridade. De acordo com Silveira (2010), tais medidas constituem a base desta metodologia, visto que a análise de agrupamento se inicia com a obtenção de uma matriz $n \times n$ de medidas de similaridade ou dissimilaridade obtidas da matriz $n \times p$ de dados. Sendo n o número unidades amostrais, que são medidas de acordo com p variáveis.

Quanto maior o valor das medidas de similaridade, mais semelhantes são as observações, enquanto que para medidas de dissimilaridade, quanto maior seu valor, menos semelhantes, ou mais divergentes, são as observações (JOHNSON e WICHERN, 1998). Em se tratando de medidas de similaridade, a mais utilizada é o coeficiente de correlação de Pearson, e se tratando de medidas de dissimilaridade, ou medidas de distâncias, as principais são a distância Euclidiana, distância Euclidiana média e distância de Mahalanobis (CRUZ et al., 2004).

Escolhida a medida de distância, ou de semelhança, passa-se então a escolher um método de agrupamento. Os métodos de agrupamento são frequentemente classificados em 2 tipos: métodos não-hierárquicos e métodos hierárquicos.

2.4.1- Métodos não-hierárquicos

Os métodos não-hierárquicos são métodos que têm como objetivo encontrar diretamente uma partição de n elementos em k grupos (clusters), de modo que a partição satisfaça dois requisitos básicos: semelhança interna e isolamento (ou separação) dos clusters formados. Para se buscar a melhor partição de ordem k , algum critério de qualidade de partição deve ser empregado. Os métodos das k -médias e o Fuzzy c -médias são exemplos de alguns métodos não hierárquicos mais utilizados (Mingoti, 2007).

2.4.2 – Métodos hierárquicos

Os métodos hierárquicos são os mais utilizados e se dividem em aglomerativos e divisivos. Os métodos aglomerativos iniciam com tantos grupos quanto às unidades amostrais, ou seja, cada unidade amostral inicialmente forma um agrupamento. Em seguida, as n unidades amostrais mais similares são agrupadas e fundidas formando os grupos. Assim, por meio de sucessivas fusões das n unidades, vão se formando $n-1$, $n-2$,..., $n-k$ grupos, até reunir todas unidades em um único grupo. Portanto, no estágio inicial do processo de agrupamento, cada elemento amostral é considerado como um cluster de tamanho 1 e no último estágio de agrupamento tem-se apenas um único cluster constituído de todos os elementos amostrais. Segundo Mingoti (2007), em termos de variabilidade, no estágio inicial, tem-se a partição com a menor dispersão interna possível, já que todos os grupos têm um único elemento e, logo a variância de cada um deles é igual a zero. No estágio final tem-se a maior dispersão interna possível, já que todos os elementos amostrais estão em um único cluster. Dentre os

métodos aglomerativos mais utilizados estão: o método do vizinho mais próximo, método do vizinho mais distante, método da ligação média, método do centróide e o método de Ward, proposto por Ward (1963).

Nos métodos divisivos, a idéia é partir de um único grupo contendo todos os elementos, e por meio de divisões sucessivas, obter vários outros sub-grupos, de forma que exista a máxima semelhança entre os elementos dos mesmos sub-grupos e a máxima dissimilaridade entre os elementos de subgrupos distintos. O processo pode ser repetido até que haja tantos subgrupos quantos elementos amostrais.

Os métodos hierárquicos se caracterizam então pelo estabelecimento de uma hierarquia em que as unidades amostrais são agrupadas por um processo que se repete em vários níveis. O resultado final destes agrupamentos podem ser representados por gráficos também denominados de dendrogramas. Há duas formas de se representar um dendograma: horizontal e vertical.

2.5 - Determinação do número ótimo de grupos

A análise de agrupamento constitui uma das mais importantes técnicas de classificação multivariada, uma vez que permite classificar várias unidades amostrais em um número reduzido de grupos contendo unidades semelhantes. Porém, uma das grandes lacunas desta metodologia, quando utilizados os métodos hierárquicos, é a determinação do número ótimo de grupos. Segundo Faria (2009) não há um critério objetivo que determine o ponto de corte do dendrograma.

Mojena (1977) propôs um critério baseado na maior amplitude das distâncias de junção dos grupos formados com o objetivo de determinar um número de grupos k que otimiza a qualidade do agrupamento dos dados. Sharma (1996) apresenta estatísticas que permitem acessar o número ótimo de grupos, e dentre estas destacam-se os índices RS (*R-square*) e RMSSTD (*Root Mean Square Standart Deviation*), sendo este último de maior utilização na área de Biometria (CECON et al., 2008; ARAÚJO, 2008; FARIA, 2009; FIORINI, 2008).

Em resumo, segundo Faria (2009), o índice RMSSTD é usado para calcular a homogeneidade dos agrupamentos, de forma que quanto mais compactos forem os grupos formados, situação esta verificada na presença de um grande número de grupos, menores os seus valores. Assim, é possível confeccionar um gráfico que mostra o decréscimo do RMSSTD em função do aumento do número de grupos. No entanto, esta trajetória não é linear, e o seu ponto de máxima curvatura indica um limiar entre uma fase de decréscimo e uma fase de estabilização. Após este ponto, denominado de ótimo, mesmo aumentando o número de clusters, não se verifica grandes declínios nos valores do RMSSTD, ou seja, esse ponto ótimo representa o número de grupos a partir do qual a estatística mencionadas não se altera mais. Assim, mesmo aumentando o número de cluster, a precisão é a mesma.

Em relação à utilização do índice RMSSTD em recentes trabalhos na área de Biometria, Cecon et al. (2008) agruparam curvas de produção estimadas de café Conilon (*Coffea canephora*) para diferentes clones e determinaram o número ótimo de grupos por meio do RMSSTD via técnica da máxima curvatura. Fiorini (2008) em um estudo sobre curvas de progresso de requeima de genótipos de tomateiro, também agrupou as curvas estimadas por essa mesma metodologia. Faria (2009), ao avaliar métodos para determinação do número ótimo de clusters em estudo de divergência genética entre 49 acessos de pimenta (*Capsicum chinense*), recomendou o método da máxima curvatura por meio do índice RMSSTD em relação ao RS, em razão da sua maior aplicação prática.

3 – Material e Métodos

3.1 – Dados

Os dados utilizados neste trabalho foram oriundos de 410 animais machos e fêmeas da raça Nelore tradicional, nascidos entre os anos de 2001 a 2004 em dois rebanhos da região de Teófilo Otoni-MG. Estes dados fazem parte dos arquivos de registros da Associação Brasileira de Criadores de Zebu - ABCZ, com sede em Uberaba (MG). Eles se referem a pesagens teoricamente trimestrais (do nascimento aos 550 dias), tendo em vista que elas não ocorriam exatamente a cada três meses. Para que o ajuste das funções de crescimento não fosse comprometido por irregularidades das pesagens dos dados, animais com menos de sete pesagens foram excluídos do estudo.

3.2 - Estimação dos parâmetros dos modelos não-lineares

Os modelos de regressão não-linear ajustados foram os apresentados na Tabela 1, aplicados aos dados de crescimento de bovinos Nelore. Utilizou-se o método dos quadrados mínimos ordinários cujas soluções foram obtidas por meio do processo iterativo de Gauss-Newton.

3.2.1 – Método dos mínimos quadrados ordinários

Gallant (1987), citado por Mendes (2008), afirma que o método dos quadrados mínimos é utilizado na estimação dos parâmetros em modelos não-lineares, da mesma maneira que em modelos lineares. Seja uma equação de regressão não-linear

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, \dots, n.$$

Assumindo $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$ e $\varepsilon_i \sim N(0, \sigma^2)$, esta função pode ser reescrita na forma matricial

$$y_i = f(\mathbf{x}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad f(\mathbf{x}, \boldsymbol{\theta}) = f(\boldsymbol{\theta}) = \begin{bmatrix} f(\mathbf{x}, \theta_1) \\ f(\mathbf{x}, \theta_2) \\ \vdots \\ f(\mathbf{x}, \theta_n) \end{bmatrix} \quad \text{e} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

A soma dos quadrados dos erros aleatórios (SQE) deverá ser minimizada em relação a $\boldsymbol{\theta}$. Portanto a função de mínimos quadrados para um modelo não-linear é dada por:

$$\text{SQE}(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(x_i, \boldsymbol{\theta})]^2,$$

a qual pode ser representada matricialmente por:

$$\text{SQE}(\boldsymbol{\theta}) = [\mathbf{y} - f(\boldsymbol{\theta})]'[\mathbf{y} - f(\boldsymbol{\theta})]$$

Derivando $\text{SQE}(\boldsymbol{\theta})$ em relação a $\boldsymbol{\theta}$ e posteriormente igualando a equação obtida a zero tem-se o estimador de mínimos quadrados de $\boldsymbol{\theta}$ ($\hat{\boldsymbol{\theta}}$) dado por:

$$\frac{\partial \text{SQE}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial}{\partial \boldsymbol{\theta}'} [\mathbf{y} - f(\boldsymbol{\theta})]'[\mathbf{y} - f(\boldsymbol{\theta})] = -2[\mathbf{y} - f(\boldsymbol{\theta})]F(\boldsymbol{\theta})$$

$$\text{Em que: } F(\theta) = \frac{\partial f(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{\partial f_1(\theta)}{\partial \theta_1} & \frac{\partial f_1(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_1(\theta)}{\partial \theta_p} \\ \frac{\partial f_2(\theta)}{\partial \theta_1} & \frac{\partial f_2(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_2(\theta)}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\theta)}{\partial \theta_1} & \frac{\partial f_n(\theta)}{\partial \theta_2} & \dots & \frac{\partial f_n(\theta)}{\partial \theta_p} \end{bmatrix}$$

Assim, fazendo $\left. \frac{\partial \text{SQE}(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = \emptyset$, tem-se:

$$F'(\hat{\theta})[y - f(\hat{\theta})] = \emptyset$$

Portanto, o sistema de equações normais não-linear (SEN) é dado por:

$$\begin{bmatrix} \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_1} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_1} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_1} \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_2} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_2} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\hat{\theta})}{\partial \hat{\theta}_p} & \frac{\partial f_2(\hat{\theta})}{\partial \hat{\theta}_p} & \dots & \frac{\partial f_n(\hat{\theta})}{\partial \hat{\theta}_p} \end{bmatrix} \cdot \begin{pmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} f_1(\hat{\theta}) \\ f_2(\hat{\theta}) \\ \vdots \\ f_n(\hat{\theta}) \end{bmatrix} \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Não existe uma solução explícita para o SEN não-linear dado acima, então as soluções são obtidas por meio de métodos iterativos. Um dos métodos iterativos mais utilizados é a linearização da função não-linear também chamada de Método de Gauss-Newton.

3.2.2. Método de Gauss Newton

Considerando o modelo não-linear $y_i = f(x_i, \theta) + \varepsilon_i$, e $\hat{\theta}_0$ um valor tal que $F'(\hat{\theta}_0)[y - f(\hat{\theta}_0)] \approx \emptyset$. Aproximando $f(\hat{\theta})$ pelo ponto $\hat{\theta}_0$ por uma expansão em série de Taylor de 1ª ordem, tem-se:

$$f(\hat{\theta}) \approx f(\hat{\theta}_0) + F(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0) \quad (1)$$

$$F'(\hat{\theta})[y - f(\hat{\theta})] \approx \emptyset \quad (2)$$

Aplicando (1) em (2): $F'(\hat{\theta})[y - f(\hat{\theta}_0) - F(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0)] \approx \emptyset$ e multiplicando à esquerda, ambos os lados da igualdade, por $[F'(\hat{\theta})]^{-1}$, obtém-se:

$$\mathbf{y} - f(\hat{\boldsymbol{\theta}}_0) - F(\hat{\boldsymbol{\theta}}_0)\hat{\boldsymbol{\theta}} + F(\hat{\boldsymbol{\theta}}_0)\hat{\boldsymbol{\theta}}_0 \approx \boldsymbol{\phi}.$$

Logo, $F(\hat{\boldsymbol{\theta}}_0)\hat{\boldsymbol{\theta}} \approx F(\hat{\boldsymbol{\theta}}_0)\hat{\boldsymbol{\theta}}_0 + [\mathbf{y} - f(\hat{\boldsymbol{\theta}}_0)]$. Multiplicando novamente à esquerda, ambos os lados da igualdade, por $[F'(\hat{\boldsymbol{\theta}}_0)]^{-1}$, verifica-se que $\hat{\boldsymbol{\theta}} \approx \hat{\boldsymbol{\theta}}_0 + [F'(\hat{\boldsymbol{\theta}}_0)]^{-1}[\mathbf{y} - f(\hat{\boldsymbol{\theta}}_0)]$.

Fazendo $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{k+1}$ e $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_k$, tem-se para a k-ésima iteração, a expressão (3), a qual representa o processo iterativo conhecido como Gauss-Newton:

$$\hat{\boldsymbol{\theta}}_{k+1} \approx \hat{\boldsymbol{\theta}}_k + [F'(\hat{\boldsymbol{\theta}}_k)]^{-1}[\mathbf{y} - f(\hat{\boldsymbol{\theta}}_k)] \quad (3)$$

Este processo iterativo prossegue até que algum critério adotado para convergência seja atingido. No presente trabalho adotou-se o seguinte critério: quando o máximo de $d_j < \delta$, sendo, $d_j = \left| \frac{(\hat{\boldsymbol{\theta}}_{j,k+1} - \hat{\boldsymbol{\theta}}_{j,k})}{\hat{\boldsymbol{\theta}}_{j,k}} \right|$ para $j=1, 2, \dots, p$, interrompe-se o processo. O valor de δ foi especificado de acordo com a facilidade de convergência de cada modelo estudado.

3.3. Avaliadores da qualidade de ajuste

Os avaliadores da qualidade de ajuste de modelos de regressão não-lineares utilizados no trabalho foram:

a) Porcentagem de convergência (C%)

É usado quando se tem ajustes individuais, ou seja, o interesse reside em ajustar uma curva para dados de cada animal. Sendo assim, é possível observar qual modelo apresenta maior facilidade de convergência, que é dado pela porcentagem de ajustes que convergiram.

b) Quadrado médio do erro

Representa a estimativa da variância residual. Quanto menor o seu valor, melhor será considerado o modelo. É dado por:

$$QME = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n - p)}$$

em que:

n é o número de observações utilizadas para ajustar a curva

p é o número de parâmetros na função

c) Coeficiente de determinação (R^2)

Este coeficiente representa a medida da capacidade do modelo de se ajustar as observações. A tendência do R^2 é aumentar de acordo com o número de parâmetros incluídos no modelo, ou seja, quanto mais parâmetros no modelo maior é o valor de R^2 . Assim, em modelos muito parametrizados, não deve-se adotar esse coeficiente como único avaliador de qualidade de ajuste. Mas de forma geral, deve-se adotar o critério, de quanto maior o R^2 melhor o modelo. (Silva, 2005). Sua expressão é dada por:

$$R^2 = 1 - \frac{SQR}{SQT}$$

SQR é a soma de quadrado de resíduos do modelo
 SQT é a soma de quadrados total corrigida

d) critério de Akaike (AIC)

Permite utilizar o princípio da parcimônia na escolha do melhor modelo, ou seja, de acordo com este critério nem sempre o modelo mais parametrizado é melhor (BURNHAM e ANDERSON, 2004). O AIC é usado para comparar modelos não-aninhados ou quando estão sendo comparados três ou mais modelos. Menores valores de AIC refletem um melhor ajuste (AKAIKE, 1974). Sua expressão é dada por:

$$AIC = -2\loglike + 2p$$

em que: p é o número de parâmetros e \loglike o valor do logaritmo da função de verossimilhança considerando as estimativas dos parâmetros.

e) critério de informação Bayesino (BIC)

Assim como o AIC, também leva em conta o grau de parametrização do modelo, e da mesma forma, quanto menor for o valor de BIC (SCHWARZ, 1978), melhor será o ajuste do modelo. É um critério assintótico, cuja adequação está fortemente relacionada com a magnitude do tamanho de amostra. Em relação a penalização aplicada na quantidade de parâmetros, esta será mais pesada do que a do AIC para amostras pequenas. Sua expressão é dada por:

$$BIC = -2\loglike + p\ln(n)$$

em que:
 n é o número de observações utilizadas para ajustar a curva.

f) Erro quadrático médio de predição (MEP)

Um dos critérios mais eficientes para testar a qualidade de um modelo de regressão é o erro quadrático médio de predição (CHROBOK et al., 2004), dado por:

$$MEP = \frac{\sum_{i=1}^n (y_i - \hat{y}_i^*)^2}{n},$$

em que: \hat{y}_i^* representa os dados estimados considerando valores dos parâmetros provenientes de um ajuste sem a presença da observação i e n é o número de observações.

g) Coeficiente de Determinação de Predição (R_p^2)

Este avaliador é equivalente ao R^2 , apresentando a mesma interpretação. Porém a diferença é que as somas de quadrado do erro e total são substituídas, respectivamente, pelos termos $nMEP$ e $\sum_{i=1}^n \hat{y}_i^{*2} - n\bar{y}^2$, os quais contemplam a influência da ausência de cada observação no ajuste do modelo, sendo:

$$\hat{R}_p^2 = 1 - \frac{nMEP}{\sum_{i=1}^n \hat{y}_i^{*2} - n\bar{y}^2},$$

em que: \bar{y} é a média dos n valores y_i , $i=1,2,\dots,n$.

Após a obtenção dos avaliadores apresentados para cada um dos modelos estudados, confeccionou-se um conjunto de dados multivariados, no qual os modelos correspondem às unidades, ou aos indivíduos, e os avaliadores às variáveis. Dessa forma, torna-se possível empregar a análise de agrupamento na tentativa de identificar modelos cujos avaliadores são semelhantes.

Vale a pena ressaltar, que ao considerar o conjunto de dados individuais, para cada modelo, os valores dos avaliadores que constituem o conjunto de dados multivariados são dados pelas médias dos valores dos avaliadores obtidos para cada animal. Para calcular estas médias foram confeccionados gráficos box-plot com o intuito de identificar possíveis *outliers*, os quais poderiam vir a comprometer a classificação dos modelos, sendo assim, uma vez identificados tais valores, os mesmos foram removidos.

3.4 - Análise de Agrupamento

O método de agrupamento utilizado pelo SAS[®] (SAS, 2003) é o método do centróide. Neste método, a distância entre dois grupos é definida como sendo a distância entre os vetores de médias, também chamados de centróides, dos grupos que estão sendo comparados. Segundo (Mingoti, 2007), essa distância é dada então pela distância euclidiana ao quadrado entre os centróides dos dois grupos.

Assim, segundo Silveira (2010), se R e S são os dois grupos a serem comparados, sendo que R possui n_r indivíduos, ou unidades, e S, n_s indivíduos. Então os centróides dos grupos são:

$$\bar{y}_r = \frac{\sum_{r=1}^{n_r} y_r}{n_r} = \begin{bmatrix} \bar{y}_{r1} \\ \bar{y}_{r2} \\ \vdots \\ \bar{y}_{rp} \end{bmatrix} \quad \text{e} \quad \bar{y}_s = \frac{\sum_{s=1}^{n_s} y_s}{n_s} = \begin{bmatrix} \bar{y}_{s1} \\ \bar{y}_{s2} \\ \vdots \\ \bar{y}_{sp} \end{bmatrix},$$

em que y_r e y_s são os vetores de observações dos objetos pertencentes aos grupos R e S, respectivamente.

A partir dos centróides, devemos obter a distância euclidiana quadrática entre os dois grupos por

$$d_{rs}^2 = \sum_{u=1}^p (\bar{y}_{ru} - \bar{y}_{su})^2. \quad (4)$$

Segundo o autor, o processo começa com cada indivíduo sendo um grupo, onde o seu centróide é um vetor de observações p -variado. Calcula-se a distância euclidiana quadrática entre todos os pares e obtemos a matriz de dissimilaridades $D = [d_{ij}^2] (n \times n)$. Quando um novo grupo (RS) é formado a partir da junção dos grupos R e S, de distância quadrática mínima entre todos os pares de grupos, o seu centróide é dado pela média ponderada

$$\bar{y}_{rs} = \frac{n_r \bar{y}_r + n_s \bar{y}_s}{n_r + n_s}.$$

Após a união dos grupos R e S, deve-se determinar a distância entre o grupo RS e os demais grupos. Assim, para um grupo T qualquer, a distância quadrática $d_{(rs)t}^2$ pode ser dada pela fórmula (4), a partir dos centróides dos grupos RS e T. No entanto, se temos apenas a matriz de dissimilaridades, pode-se utilizar uma fórmula de atualização de distâncias quadráticas, dada por

$$d_{(rs)t}^2 = \frac{n_r}{n_r+n_s} d_{rt}^2 + \frac{n_s}{n_r+n_s} d_{st}^2 - \frac{n_r n_s}{(n_r+n_s)^2} d_{rs}^2 \quad (5)$$

em que as distâncias quadráticas d_{rt}^2 , d_{st}^2 e d_{rs}^2 são provenientes da matriz de dissimilaridades antes da atualização efetuada em função do agrupamento dos grupos R e S.

Utilizando-se a equação (5) pode-se atualizar a matriz de distâncias após o agrupamento dos grupos R e S, e então, aplicar o algoritmo geral dos métodos hierárquicos para esse método a partir da matriz de dissimilaridades, sem a necessidade de utilizar os dados originais (FERREIRA, 2008). A utilização deste algoritmo permite obter número máximo de grupos correspondente ao número de indivíduos, ou unidades, e a visualização de todos estes possíveis grupos torna-se possível mediante a construção do dendograma. Dessa forma, cabe ao pesquisador identificar nesta figura o número adequado, ou ótimo, de grupos. Como muitas vezes tal identificação é realizada de forma subjetiva, no presente trabalho será apresentada uma técnica estatística baseada no cálculo do RMSSTD com o intuito de eliminar a subjetividade.

3.5 - Determinação do número ótimo de grupos e escolha do melhor modelo

O índice RMSSTD mede a homogeneidade dos agrupamentos, ou seja, quanto menor o RMSSTD mais homogêneos serão os grupos, condição esta verificada na presença de um grande número de grupos (Silveira, 2010). Este índice segundo esse autor é dado por:

$$\text{RMSSTD}_k = \sqrt{\frac{SQ_1 + SQ_2 + \dots + SQ_p}{gl_1 + gl_2 + \dots + gl_p}},$$

em que: $SQ_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ é a soma de quadrado da j-ésima variável calculada, considerando as n observações presentes em cada novo cluster k, ou seja, a cada novo cluster obtém-se um novo valor para o índice em questão.

Após calcular os valores do RMSSTD, é possível construir um gráfico que representa o comportamento desse índice em função do número de grupos (Nc), como ilustrado na Figura 1(a). Este comportamento pode ser descrito por um modelo exponencial simples, como $\text{RMSSTD} = a(Nc)^{-b}$, em que a e b são os parâmetros deste modelo. Assim, o ponto de máxima curvatura desta curva, o qual corresponde ao número ótimo de grupos (Xo), pode ser determinado geometricamente (LARSON et al., 1998) por meio da interseção desta curva com uma reta, de forma que a maior distância entre elas corresponda ao ponto em questão, como mostrado na Figura 1 (b). Mas é importante ressaltar, que esse método é eficiente somente quando a qualidade do ajuste do modelo exponencial é alto, ou seja, quando encontramos altos valores para R^2 . Quanto pior for o ajuste do modelo exponencial menos confiável será o emprego desse método.

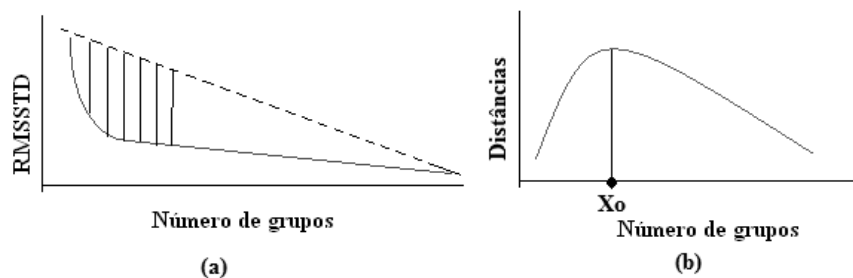


Figura 1. Ilustração da determinação geométrica do ponto de máxima curvatura da curva representativa do comportamento do RMSSTD em função do número de grupos

Após a obtenção dos grupos (*clusters*), realizou-se uma ordenação destes de acordo com as médias calculadas para os avaliadores dos modelos presentes dentro de cada um deles. Como critérios de ordenação, foram considerados os maiores valores de R^2 , R^2_p e percentual de convergência e menores valores de AIC, BIC e MEP.

3.6 - Aspectos computacionais

Os modelos de regressão não-linear (Tabela 1) foram ajustados aos dados individuais de crescimento de bovinos por meio do PROC MODEL do SAS[®] (SAS, 2003). Como relatado, utilizou-se o método dos quadrados mínimos ordinários via método iterativo de Gauss-Newton.

Em relação aos avaliadores da qualidade de ajuste (item 3.3) cabe ressaltar que os valores do R^2_{aj} e a *loglike*, a qual foi usada para obter os valores do AIC e BIC, foram obtidos diretamente do PROC MODEL. Para a obtenção do MEP e do R^2_p , elaborou-se uma adaptação na rotina deste procedimento para remover uma observação a cada ajuste. A porcentagem de convergência diz respeito aos animais cuja convergência foi observada até um número máximo de 1000 iterações.

Após a obtenção dos avaliadores de qualidade para cada modelo, confeccionaram-se então conjuntos de dados multivariados como descrito no item 3.3, os quais foram submetidos à análise de agrupamento no PROC CLUSTER considerando o método centróide. Nesta análise obtiveram-se os valores dos RMSSTD em relação ao número de clusters, a qual gerou o gráfico que disponibilizou a identificação do número ótimo de cluster tendo em vista o método da máxima curvatura apresentado no item 3.5. Como complemento dessa última análise, utilizou-se o PROC TREE para visualizar o dendograma e para verificar quais modelos pertencem aos diferentes grupos obtidos pela discriminação estatística em relação aos avaliadores de qualidade.

4-Resultados e Discussão

Na tabela 2 estão apresentadas as médias das estimativas dos parâmetros dos modelos de crescimento ajustados aos dados individuais de bovinos Nelore.

TABELA 2 - Estimativas dos parâmetros para os modelos de crescimento de bovinos Nelore considerando o ajuste individual

Modelos	Estimativas			
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Schnute	272,6673	213,9545*	-	135,3977
Mitscherlich	495,1949	-35,5899	0,002412	-
Richards	173,9015	138,6617	0,865346	-7,4132
Gompertz	383,0079	0,8123	0,006421	-
Logístico	343,2097	1,7838	0,009536	-
Meloun I	231,4542	179,7629	0,822111	-
Meloun II	468,9182	-6,0151	0,002625	-
Brody	495,3454	0,9289	0,002434	-
von Bertalanffy	410,5213	0,5415	0,004652	-
Michaelis-Menten	553,6187	418,8164	-	-
Michaelis-Menten Modificado	124,8109	133251,8	424,9587	-37,9608
Weibull	103,0795	-3543,75	312,1085	-56,9986

$\hat{\beta}_1$ representa o peso adulto estimado, $\hat{\beta}_2$ representa a constante de integração estimada, $\hat{\beta}_3$ representa o taxa de maturação estimada e $\hat{\beta}_4$ representa o ponto de inflexão estimado.* Por motivos de dificuldade de convergência adotou-se uma reparametrização na qual o valor 213,9545 representa o produto $\beta_2\beta_3$.

A partir dos resultados apresentados na tabela 2, percebe-se que os modelos Richards, Michaelis-Menten Modificado e Weibull subestimaram o peso adulto (β_1), sendo que os outros modelos estimaram valores para este parâmetro condizentes com a literatura especializada (Tedeschi, 1996; Carrijo et al., 1999 ; Garnero et al., 2005, Santoro et al., 2005). Para o parâmetro β_3 , que representa a taxa de maturidade, com exceção de Michaelis-Menten Modificado e Weibull, percebe-se uma grande amplitude de variação da estimativa deste parâmetro entre os outros modelos avaliados (0,0002 – 0,8653). De acordo com Silva et al. (2004) essa grande variação para este parâmetro pode ser atribuída à alta e negativa correlação entre a taxa de maturidade e o peso adulto, assim os modelos que apresentam menores valores para $\hat{\beta}_1$ apresentam altos valores para $\hat{\beta}_3$. De forma geral, optou-se por discutir o valor do peso adulto (β_1) devido a sua maior interpretação biológica e também por este parâmetro ser mais relatado na literatura que os demais.

Na tabela 3 estão apresentados os resultados dos avaliadores da qualidade de ajuste para cada modelo, considerando os dados individuais de crescimento.

Tabela 3 - Resultados dos avaliadores da qualidade de ajuste para os modelos não lineares considerando ajuste individual.

Avaliadores da Qualidade de Ajuste							
Modelo	R ²	QME	AIC	BIC	MEP	R ² _p	C%
A - Schnute	0,7322	5283,16	78,7738	78,6115	7234,80	0,9150	100
B - Mitscherlich	0,9747	372,53	61,2488	61,0866	868,042	0,9880	79,9517
C - Richards	0,7189	7150,50	79,1788	78,9624	4693,33	0,9195	8,4541
D - Gompertz	0,9711	457,17	63,2762	63,1142	1067,22	0,9859	97,1014
E - Logístico	0,9646	579,71	65,3381	65,1759	1304,14	0,9835	98,7923
F - Meloun I	0,9672	493,63	61,7026	61,5403	1221,06	0,9834	78,744
H - Meloun II	0,9710	447,82	62,3541	62,1918	6868,61	0,8929	37,43961
N - Brody	0,9741	381,86	61,3352	61,1729	763,22	0,9894	80,6763
O - von Bertalanffy	0,9730	420,82	62,4535	62,2913	784,87	0,9895	95,8937
P - Michaelis-Menten	0,9661	570,06	65,3135	65,2048	862,102	0,9877	85,2657
Q - Michaelis-Menten M.	0,7318	3976,17	65,0316	64,8153	2576,92	0,9537	1,93237
R - Weibull	0,7640	2780,31	74,6309	74,4146	5590,13	0,9227	12,07729
Média	0,8923	1909,47	66,7197	66,5484	2819,537	0,9592	64,6940
Desvio-padrão	0,1154	2347,22	6,7626	6,7480	2541,761	0,0364	38,3227
CV%	12,9364	122,92	10,1358	10,1400	90,148	3,8042	59,2369

R² - coeficiente de determinação, QME - Quadrado médio do erro, AIC- valor do critério de Akaike, BIC-valor do critério de informação Bayesiano, MEP-erro quadrático médio de predição, R²_p - coeficiente de determinação de predição, C%-porcentagem de convergência

Observa-se na tabela 3 que para praticamente todos os avaliadores de qualidade, com exceção do QME, MEP e C%, apresentaram baixo coeficiente de variação (CV%). Resultados análogos foram observados por Silveira et al, (2010) em seu trabalho com ovinos. A grande amplitude observada para C% deve-se ao baixo número de animais que convergiram nos modelos com quatro parâmetros, Richards, Michaelis-Menten Modificado e Weibull, mostrando então a dificuldade de que se tem em trabalhar com modelos mais parametrizados. Em relação ao QME, os maiores valores também foram observados para os modelos mais parametrizados. Em relação ao alto CV obtido para o MEP, isso se deveu-se provavelmente porque este avaliador é calculado com base na remoção sucessiva de observações dos conjuntos de dados originais, o que resulta este avaliador a representar a capacidade dos modelos em se ajustarem bem na presença de dados faltantes. Mas segundo Nobre et al. (1987) a qualidade dos modelos varia significativamente com a perda de observações, assim a grande variação do MEP se deve a perda de qualidade de alguns destes modelos nestas condições. Em relação aos outros avaliadores de qualidade, pode-se perceber que de forma geral, os modelos da família de Richards foram os que apresentaram os melhores avaliadores de qualidade, mostrando a superioridade destes modelos em se trabalhar com dados de bovinos.

Os resultados apresentados na tabela 3 de avaliadores da qualidade de ajuste foram os dados utilizados como conjunto de dados multivariados, para a realização da análise de agrupamento.

Para a determinação do número ótimo de cluster foi utilizada a metodologia RMSSTD. Primeiro foi obtido os valores do índice RMSSTD, e estes valores foram plotados em um gráfico de acordo com o número de clusters como é mostrado na figura 2. Ainda, nesta figura pode-se observar a equação exponencial ajustada para este índice de acordo com o número de clusters e o valor do R².

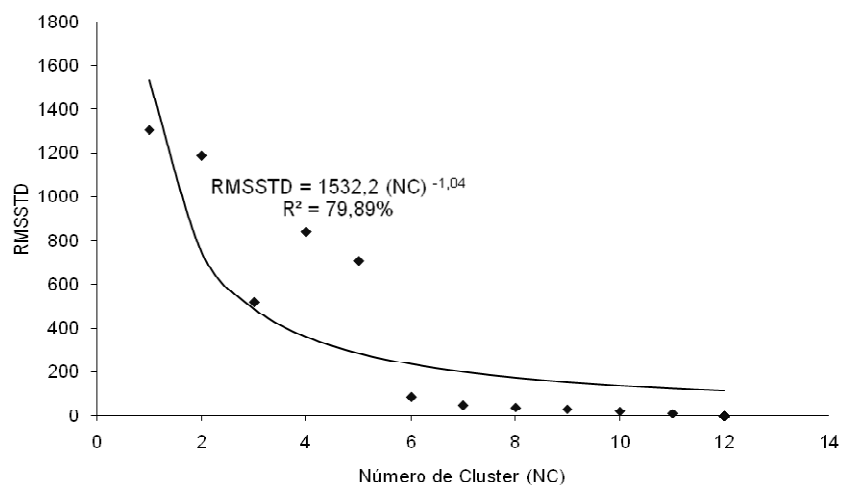


Figura 2 – Ajuste exponencial do índice RMSSTD em função do número de cluster

Observa-se pela figura 2, que o valor do R^2 foi abaixo de 80%, indicando que o ajuste não foi tão bom. Assim, o uso do método de máxima curvatura não é o mais indicado. Mas por inspeção visual, pode-se notar, que a partir de 6 grupos tem-se clara estabilização da curva, sendo então este o número considerado como ótimo para o número de clusters. Os grupos formados podem ser visualizados no dendrograma da figura 3.

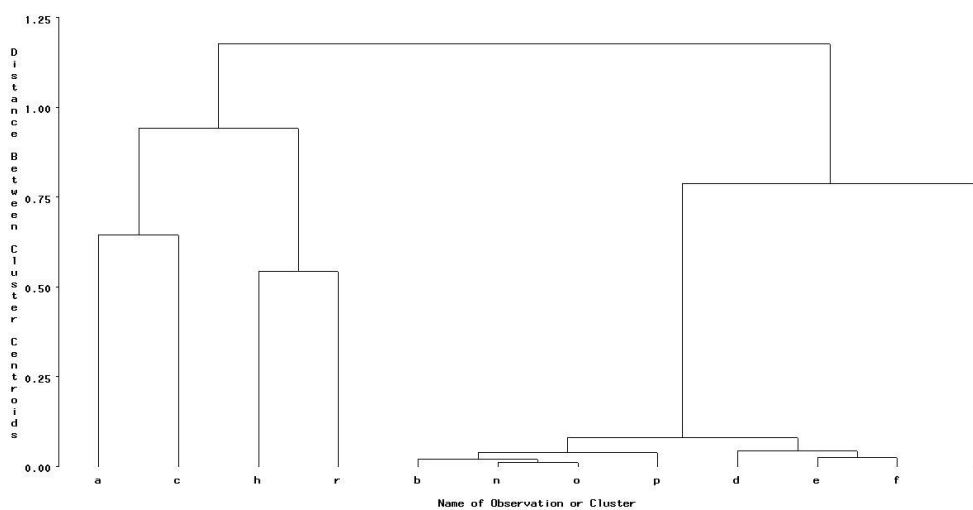


Figura 3 – Dendrograma resultante da análise de agrupamento.

Na Tabela 4 são apresentados os *clusters* com os seus respectivos modelos e médias dos valores dos avaliadores de qualidade do ajuste.

Tabela 4 - *Clusters* com os seus respectivos modelos e médias dos avaliadores

Avaliador	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 4</i>	<i>Cluster 6</i>	<i>Cluster 3</i>	<i>Cluster 5</i>
	Modelos {N,O,B,E,F,P,D}	Modelos {H}	Modelos {A}	Modelos {Q}	Modelos {R}	Modelos {C}
R^2_{aj}	0,9701	0,971	0,7322	0,7318	0,764	0,7189
QME	467,90	447,81	5283,16	3976,17	2780,31	7150,55
AIC	62,9525	62,3541	78,7738	65,0316	74,6309	79,1788
BIC	62,7979	62,1918	78,6115	64,8152	74,4146	78,9624
MEP	981,5252	6868,61	7234,8	2576,92	5590,13	4693,33
R^2_p	0,9868	0,8929	0,915	0,9537	0,9227	0,9195
% C	88,0607	37,43961	100	1,9323	12,07729	8,4541

Na ordenação dos clusters (tabela 4), foi dado um peso alto ao percentual de convergência (%C), uma vez que este critério é muito importante em estudos de curvas de crescimento individuais para fins de melhoramento genético. Isto porque, se um modelo apresenta qualidade de ajuste muito alta, mas percentual de convergência muito baixo, o melhorista terá um número reduzido de animais. Serão utilizados então somente aqueles animais na qual a convergência foi verificada, para serem utilizados em programas de seleção para eficiência de crescimento, medida pelas estimativas dos parâmetros β_1 e β_3 .

Ainda de acordo com os resultados apresentados na tabela 4, pode-se perceber que o melhor cluster foi de número 1, pois este foi o que apresentou de forma geral bons valores para todos os avaliadores considerados simultaneamente. Entre os sete modelos pertencentes a este grupo, pode-se notar que quatro são modelos da família de Richards. O único desta família que não está presente neste grupo superior é o modelo de Richards. Este fato confirma a boa capacidade de descrição destes quatro modelos a dados de bovinos e mostra também a exceção do modelo de Richards por sua dificuldade computacional de ajuste.

Assim, pode-se indicar entre os modelos presentes no cluster 1, e entre os da família de Richards, o Modelo de Brody como o melhor modelo para descrição do crescimento de bovinos Nelore, visto que este modelo apresentou bons avaliadores de qualidade de ajuste e parâmetros com interpretação biológica condizentes com a realidade. Este resultado pode ser confirmado por Forni (2007) que afirma que a equação de Brody é, frequentemente, citada como a mais utilizada para descrever dados de crescimento de bovinos de corte, pela facilidade de convergência dos parâmetros, mesmo na presença de dados faltantes, menor esforço computacional e qualidade de ajuste. Garnero et al. (2005) em seu trabalho, também encontrou o modelo de Brody como o mais indicado para o descrever o crescimento de animais Nelore. Além disso, em suas revisões destaca outros autores (Brown et al., 1976; Perotto et al., 1992; Oliveira et al., 2000) que encontraram a curva de Brody como indicada para descrever o crescimento tanto de bovinos Nelore quanto de outras raças.

5 - Conclusões

A metodologia de agrupamento proposta no presente trabalho mostrou-se adequada para classificar modelos não lineares tendo em vista que possibilitou a escolha do melhor modelo para descrever o crescimento de bovinos Nelore.

O modelo de Brody foi o que apresentou o melhor ajuste para o conjunto de dados de bovinos Nelore. Este modelo apresentou predição para o peso adulto condizente com a realidade, e bons avaliadores de qualidade de ajuste inclusive para o percentual de convergência sendo, portanto recomendado para o estudo de crescimento de bovinos em programas de melhoramento genético.

6 - Referências Bibliográficas

AKAIKE, H. A new look at the statistical model identification. **IEEE Transaction on Automatic Control**, v.19, p.716-723, 1974.

ARAUJO, M.N.M.de. **Análise de sobrevivência do tomateiro a *Phytophthora infestans***. 2008. 53f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2008.

BALDWIN, R. L. **Modeling ruminant digestion and metabolism**. London (UK): Chapman & Hall; 1995.

BRODY, S. **Bioenergetics and Growth**. Rheinhold Publishing, New York. 1945.

BROWN, J.E.; FITZHUGH Jr., H.A.; CARTWRIGHT, T.C. A comparison of nonlinear models for describing weight-age relationships in cattle. **Journal of Animal Science**, Champaign, v.42, n.4, p.810-818, Apr. 1976.

BURNHAM, K.P.; ANDERSON, D.R. Multimodel Inference: Understanding AIC and BIC in Model Selection. **Sociological methods & research**, v.33, n.2, p.261-304, Nov. 2004.

CARRIJO, S.M.; DUARTE, F.A.M. Description and comparison of growth parameters in Chianina and Nelore cattle breeds. **Genetics and Molecular Biology**, São Paulo, v.22, n.2, p.187-196, jun. 1999.

CECON, P.R.; SILVA, F.F.; FERREIRA, A.; FERRÃO, R.G.; CARNEIRO, A.P.S.; DETMANN, E.; FARIA, P.N.; MORAIS, T.S.da.S. Análise de medidas repetidas na avaliação de clones de café 'Conilon'. **Pesquisa Agropecuária Brasileira**, Brasília, v.43, n.9, p.1171-1176, set. 2008.

CHROBOK, V.; MELOUN, M.; SIMÁKOVÁ, E. Descriptive growth model of the height of stapes in the fetus: a histopathological study of the temporal bone. **Eur Arch Otorhinolaryngol**, 261, p.25-29, 2004.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 3.ed.- Viçosa: UFV, 480p. 2004. vol 1

ELIAS, A.M. **Análise de curvas de crescimento de vacas das raças Nelore, Guzerá e Gir**. 1998. 128f. Dissertação (Mestrado em Ciência Animal e Pastagens)-Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, 1998.

FARIA, P.N. **Avaliação de métodos para determinação do número ótimo de clusters em estudo de divergência genética entre acessos de pimenta**. 2009. 54f. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa, MG, 2009.

FERREIRA, D.F. **Estatística Multivariada**. 1.ed. – Lavras: Ed. UFLA, 662p. 2008.

FERREIRA FILHO, P. **Introdução ao planejamento e análise estatística de experimentos**. Universidade Federal de São Carlos, 2009, 10 p.

FIORINI, C.V.A. **Introgressão de genes de resistência à requeima de *Solanum habrochaites* em *Solanum lycopersicon***, 2008. Tese (Doutorado em Fitotecnia - Produção Vegetal) – Universidade Federal de Viçosa, Viçosa, MG, 2008.

FITZHUGH Jr., H.A. Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, Champaign, v.42, n.4, p.1036-1051, Apr.1976.

FORNI, S. **Análise da curva de crescimento de bovinos da raça nelore utilizando funções não lineares em análises bayesianas** 2007. 75p. Tese (Doutorado em Zootecnia)-Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal.

FREITAS, A.R.de. Curvas de crescimento na produção animal. **Revista Brasileira de Zootecnia**, v.34, n.3, p.786-795, 2005.

GALLANT, A. R. **Nonlinear statistical models**. New York: John Wiley, 1987. 610 p.

GARNERO, A. V.; MARCONDES, C. R.; BEZERRA, L. A. F.; OLIVEIRA, H. N.; LÔBO, H. N. Parâmetros genéticos da taxa de maturação e do peso assintótico de fêmeas da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v.57,n.5, p.652- 662,out. 2005.

GOMPERTZ, B.; PHILOS, T. Roy Soc London. 115-513. 1825.

GOUS, R. M. Making progress in the nutrition of broilers. **Poultry Science**, v.77, n.1, p.111-117, 1998.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. New Jersey-USA: Englewood Cliffs, 642p. 1992.

LARSON, R.; HOSTETLER, R.; EDWARDS, B. **Cálculo com aplicações**. 5.ed. Rio de Janeiro: LTC, 1998.

LOPEZ, S.; FRANCE, J.; GERRITS, W.J.; DHANOA, M.S.; HUMPHRIES, D.J.; DIJKSTRA, J. A generalized Michaelis-Menten equation for the analysis of growth. **Journal of Animal Science**, v.78, p.1816-1828, 2000.

MAZZINI, A. R. A.; MUNIZ, J. A.; SILVA, F. F.; AQUINO, L. H. Curvas de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, Santa Maria, v. 35, n. 2, p. 422-427, mar./abr. 2005.

MAZZINI, A. R. A.; MUNIZ, J. A.; SILVA, F. F.; AQUINO, L. H. Análise da curva de crescimento de machos Hereford. **Ciência e Agrotecnologia**, Lavras, v. 27, n. 5, p. 1105-1112, set./out. 2003.

MACHADO, E.J.; MUNIZ, J.A.; SILVA, F.F. Análise da curva de crescimento por meio da equação de Michaelis-Menten. In: 11 Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO) e 50a. Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBRAS), 2005, Londrina. Anais do 11 Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO) e 50a. Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBRAS). Londrina : UEL, 2005.

MELOUN, M.; MILITKÝ, J. **Sbírka úloh, Statistické zpracování experimentálních dat**. Pardubice: Universita Pardubice. 1996.

MENDES, P.N.; MUNIZ, J.A.; SILVA, F.F.; MAZZINI, A.R.de.A.; SILVA, N.A.M Análise da curva de crescimento difásica de fêmeas Hereford por meio da função não linear de Gompertz. **Ciência Animal Brasileira**, v. 10, n. 2, p. 454-461, abr./jun. 2009

MENDES, P. N. **Curvas de crescimento difásicas de fêmeas hereford com erros auto regressivos e heterogeneidade de variâncias**, 2007. 98p. Dissertação (Mestrado em Agronomia)-Universidade Federal de Lavras, Lavras, MG.

MICHAELIS, L.; MENTEN, M.L. Die kinetik der invertinwirkung. **Biochemische Zeitschrift**, v.49, 334-336, (1913)

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada. Uma abordagem aplicada**. 1ª edição. Belo Horizonte, Editora UFMG, 2007. 295p.

MITSCHERLICH, E.A. Das Gesetz des Pflanzenwachstums. *Landwirtsch Jahrb* 53: 167-182. 1919.

MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **Computer Journal**, London, v.20, p.359-363, 1977.

NOBRE, P.R.C.; ROSA, A.N; SILVA, A.N.; EVANGELISTA, S.R.M. Curvas de crescimento de gado Nelore ajustadas para diferentes frequências de pesagens. **Pesquisa Agropecuária Brasileira**, v.22, n 9:10, p. 1027-1037, 1987.

OLIVEIRA, H.N.; LÔBO, R.B.; PEREIRA, C.S. Comparação de modelos não-lineares para descrever o crescimento de fêmeas da raça Guzerá. **Pesquisa Agropecuária Brasileira**, Brasília, v.35, n.9, p.1843-1851, set. 2000.

PAZ, C. C. P.; PACKER, I. U.; FREITAS, A. R.; TALHARI, D. T.; REGINATO, L. C. A.; ALENCAR, M. M.; CRUZ, G. M. Ajuste de Modelos Não-Lineares em Estudos de Associação entre Polimorfismos Genéticos e Crescimento em Bovinos de Corte, **Revista Brasileira Zootecnia**, v.33, n.6, p.1416-1425, 2004 .

PEROTTO, D.; CUE, R.I.; LEE, A.J. Comparison of nonlinear functions for describing the growth curve of three genotypes of dairy cattle. **Can. J. Anim. Sci.**, v.72, p.773-782, 1992.

RATKOWSKI, D.A. **Nonlinear regression modeling: a unified practical approach**. Marcel Dekker, New York. 1983.

RICHARDS, F.J. A flexible growth function for empirical use. **Journal of Experimental Botany**, Oxford, v.10, p.290-300, 1959.

RODRIGUES, A.; SILVA, F. F.; LOBATO Jr, D.; CARNEIRO, A.P.S. Avaliação da não-linearidade de modelos de crescimento via curvatura de Bates & Watts: aplicação em cordeiros da raça Bergamácia. In: 52 RBRAS E 12 SEAGRO, 2007, Santa Maria. Anais da 52 RBRAS E 12 SEAGRO. Santa Maria: UFSM, 2007.

RONDÓN, E. O. O.; MURAKAMI, A. E.; SAKAGUTI, E. S. Modelagem Computacional para Produção e Pesquisa em Avicultura **Revista Brasileira de Ciência Avícola** v. 4 n.1 p. 199-207, 2002.

SAMPAIO, I.B.M. Modelos matemáticos na nutrição animal. In: **Reunião Anual da sociedade Brasileira de Zootecnia**, 2002, Recife, PE.

SAMPAIO, I.B.M. **Experimental designs and modelling techniques in the study of roughage degradation in the rumen and growth of ruminants**. 1988. 114p. Thesis (PhD) - University of Reading, Reading, UK.

SANDLAND, R. L. Mathematics and the growth of organisms – some historical impressions. **Mathematical Scientist**, v. 8, p. 11-30, 1983.

SANTORO, R.S.; BARBOSA, S.B.P.; BRASIL, L.H.A.; SANTOS, E.S. Estimativas de parâmetros de curvas de crescimento de bovinos zebu, criados no estado de Pernambuco. **Revista Brasileira de Zootecnia**, Viçosa, MG, v.34, n.6, p.2262-2279, nov./dez. 2005.

SAS Institute Inc. **Statistical Analysis System user's guide**. Version 9.1 ed. Cary: SAS Institute, USA, 2003.

SCHNUTE, J. **A versatile growth model with statistically stable parameters**. Can J Fish Aquat Sci 38: 1128-1140. 1981.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v.6, p.461-464, 1978.

SHARMA, S. **Applied multivariate techniques**. New York: John Wiley, 493p., 1996.

SILVA, N. A. M., LIMA, R. R., SILVA, F. F. E., AQUINO, L. H., MUNIZ, J. A. Modelo Hierárquico Bayesiano aplicado na avaliação genética de curvas de crescimento de bovinos de corte. **Arquivo Brasileiro de Medicina Veterinária e zootecnia.**, v.62, n.2, p.647-654, 2010

SILVA, M. A. **Conceitos de Análise de Dados**, Cadernos Didáticos, Editora: FEPMVZ. Escola de veterinária, UFMG, Belo horizonte, p.189, 2005

SILVA, N.A.M. da; AQUINO, L.H. de; SILVA, F.F; OLIVEIRA, A.I.G. de. Curvas de crescimento e influência de fatores não-genéticos sobre as taxas de crescimento de bovinos da raça Nelore. **Ciência e Agrotecnologia**, Lavras, v.28, n.3, p.647-654, maio./jun. 2004.

SILVEIRA, F. G. **Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte**. 2010. 55p. Dissertação (Mestrado em Estatística aplicada a Biometria)-Universidade Federal de Viçosa, Viçosa, MG.

SILVEIRA, F.G.; SILVA, F. F.; CARNEIRO, P.L.S.; MALHADO, C.H.M.; PETERNELLI, L.A.; SOUZA Jr, A.A.O. Classificação multivariada de modelos de crescimento para grupos genéticos de ovinos de corte. In: 54ª RBRAS (Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria) e 13º SEAGRO (Simpósio de Estatística Aplicada à Experimentação Agrônômica), 2009, São Carlos-SP. Anais da 54ª RBRAS e 13º SEAGRO. São Carlos-SP: UFScar, 2009.

TEDESCHI, L.O. **Determinação dos parâmetros da curva de crescimento de animais da raça Guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação**. 1996. 140p., Dissertação (Mestrado em Zootecnia)- ESALQ, Piracicaba, SP

VON BERTALANFFY, L. **Quantitative laws for metabolism and growth.** Q. Rev. Biol. 32:217–231, 1957.

WARD, J.H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, v.58, p.236-244, mar., 1963.

CAPÍTULO 2

PROPOSTA DE ANÁLISE GENÉTICA DE CURVAS DE CRESCIMENTO DE BOVINOS POR MEIO DO ALGORITMO SAEM

Resumo

A estimação dos componentes de (co)variância dos parâmetros de modelos de crescimento pode ser feita por vários métodos. A partir dos componentes estimados podem ser encontradas as herdabilidades e as correlações genéticas (parâmetros genéticos) referentes aos parâmetros do modelo. As estimativas dos parâmetros (modelo e genéticos) são podem ser utilizadas em programas de melhoramento animal. O objetivo deste capítulo foi comparar duas diferentes metodologias na avaliação genética de curvas de crescimento de animais Nelore: o algoritmo SAEM e o método Two Step. Para a implementação dessas metodologias foram utilizados o modelo de crescimento de Brody modificado e o modelo touro. A diferença marcante entre SAEM e Two Step, é que o algoritmo SAEM estima simultaneamente parâmetros do modelo e efeitos genéticos e ambientais e o método Two Step faz esse processo de estimação em duas etapas distintas. Foram obtidos a partir das metodologias testadas, além das estimativas de efeitos fixos e parâmetros genéticos, os valores genéticos preditos para os touros avaliados. A partir dos valores genéticos preditos foram obtidas as curvas genéticas para os touros. O algoritmo SAEM se mostrou consistente na estimação dos efeitos fixos e predição dos efeitos aleatórios, se apresentando como uma alternativa viável para avaliação genética de animais Nelore.

Palavras-chave: componentes de (co)variância, algoritmo SAEM , curvas genéticas

PROPOSAL FOR GENETIC ANALYSIS OF GROWTH CURVES OF CATTLE THROUGH THE SAEM ALGORITHM

Abstract

The estimation of the (co)variance components for the parameters of the growth models can be evaluated by many methods. From the estimated components can be found the genetic parameters of the parameters of model. These estimates (model and genetics) can be used in animal breeding programs. The objective of this chapter was to compare two methodologies in genetic evaluation of growth curves of Nelore cattle: the SAEM algorithm and Two Step method. To implement these methodologies were used Brody modified growth curve and the sire model. The difference between SAEM and Two Step is that SAEM estimates jointly the parameters of the model and genetics and environmental effects and the Two Step method makes this process in two independent steps. Were obtained from the methodologies estimates of the fixed effects and genetics parameters and predict breeding values for the sires. From the breeding values were obtained genetics curves for the sires. The SAEM algorithm proved consistent in the estimation of fixed effects and prediction of random effects.

Key-Words: (co)variance components, SAEM algorithm, genetics curves

1-Introdução

Em rebanhos de gado de corte, o crescimento dos animais é geralmente avaliado por um conjunto de pesagens tomadas em um mesmo indivíduo, em intervalos regulares ou não, desde o nascimento até a idade adulta. A partir dessas medidas é possível construir uma curva do peso em relação à idade que representa todo o processo de crescimento dos animais.

Modelos de regressão não linear têm se mostrado adequados para descrever curvas de crescimento, pois apresentam parâmetros que podem ser interpretados biologicamente. Estes parâmetros são caracterizados principalmente pelo peso à maturidade, que representa o peso na idade adulta, e pela velocidade de crescimento, que representa uma medida de precocidade. Em geral, ajustam-se os modelos não-lineares a um conjunto de dados de peso e idade, com a finalidade de prever a variável dependente (peso) para valores da variável independente (idades).

Além dos parâmetros que são estimados diretamente do modelo e que apresentam interpretação biológica, podem ser obtidos também os parâmetros genéticos referentes a eles, como as herdabilidades e correlações, os quais podem ser utilizados em programas de melhoramento animal.

Os valores de herdabilidade para as estimativas dos parâmetros do modelo é de moderada a alta (Blasco et al., 2003; Silva et al., 2010), indicando que estas características podem responder bem a programas de seleção. Porém, quando se dispõem de animais sujeitos a diferentes efeitos ambientais e genéticos, torna-se necessário que as estimativas dos parâmetros das curvas sejam corrigidas para tais efeitos. Essa correção, ou ajuste, é extremamente útil quando o objetivo é alterar a forma da curva de crescimento para programas de melhoramento animal que visam o aumento da eficiência do crescimento.

As estimativas dos parâmetros dos modelos de crescimento em programas de seleção, normalmente, são obtidas por meio de uma metodologia frequentista (método Two-Step), que considera duas fases distintas (Garnero et al., 2005): a primeira corresponde ao ajuste individual dos modelos não-lineares, de forma a obter um conjunto de estimativas de parâmetros para cada animal e a segunda corresponde à caracterização dessas estimativas como variáveis dependentes em um modelo misto contendo efeitos fixos e aleatórios, de forma a se obter os componentes de (co)variância necessários ao cálculo das herdabilidades e das correlações. Um fato importante decorrente da aplicação dessa metodologia é que os fatores genéticos e ambientais não são considerados no processo de estimação individual, ou seja, as estimativas dos parâmetros não são corrigidas simultaneamente para estes fatores.

De acordo com Jaffrézic et al. (2006), o algoritmo EM com o uso de um processo estocástico (algoritmo SAEM), se apresenta como uma alternativa eficiente para o estudo de crescimento de animais, pois permite estimar simultaneamente parâmetros das curvas de crescimento e dos fatores genéticos e ambientais que atuam sobre eles pelo método ML (máxima verossimilhança).

No entanto, em melhoramento genético de bovinos de corte, conhecer somente o comportamento genético dos parâmetros da curva não é uma alternativa que esteja sendo implementada na prática em programas de melhoramento de gado de corte. Assim, além de conhecimento do perfil genético de tais parâmetros, é importante conhecer a denominada curva de crescimento genética dos animais, a qual permite prever o valor genético dos animais em qualquer ponto do intervalo de tempo em que as medidas de peso idade foram tomadas, possibilitando assim realizar a seleção em qualquer idade desejada, inclusive as padrão. De forma geral, a metodologia de Regressão Aleatória permite a obtenção da referida curva genética, mas essa metodologia utiliza-se apenas de modelos lineares, como os polinômios de Legendre, cujos parâmetros não apresentam interpretação biológica como aqueles dos modelos de crescimento não lineares

Assim, mediante o que foi exposto, os objetivos da realização deste capítulo foram: Obter a partir da utilização do algoritmo SAEM e do método Two Step, os componentes de variância dos parâmetros com interpretação biológica (peso adulto, A, e taxa de maturidade, K) do modelo não linear de crescimento de Brody modificado, estimar os valores de herdabilidade de cada parâmetro do

modelo (A e K) e a correlação genética entre os parâmetros. Além disso, serão obtidas as curvas genéticas individuais, relacionadas a cada uma das metodologias empregadas, a partir dos valores genéticos preditos para os parâmetros do modelo de crescimento utilizado.

2 - Revisão de Literatura

2.1- Seleção de curvas em Bovinos

No Brasil, nos últimos anos, tem-se enfatizado a necessidade de se selecionar animais adaptados a um sistema de produção de ciclo curto, sem aumentar os custos de produção. Vários indicadores econômicos apontam nesta direção, como uma solução para viabilizar e dar poder competitivo à pecuária de corte. Com certeza, no Brasil, a enorme valorização de animais com peso adulto cada vez maior já não conquista tantos adeptos entre os criadores.

De acordo com FRIES (1995), para conseguir obter um genótipo bovino adequado a um sistema de produção de ciclo curto e sem aumentar as exigências nutricionais, é necessário alterar as formas das curvas de crescimento corporal, reduzindo ou mantendo constantes os tamanhos adultos e a idade de peso de terminação ou acabamento de carcaça.

Alterar a forma da curva de crescimento traria como vantagens: evitar o antagonismo genético entre menor velocidade de crescimento da progênie e maior tamanho e custo de manutenção de vacas; reduzir os problemas de partos distócicos por meio da diminuição da relação peso ao nascer dos bezerras/peso da vaca e diminuir a idade à primeira cria (Fitzhugh Jr, 1976).

Na maioria dos trabalhos sobre curvas de crescimento, as diferentes pesagens são utilizadas para descrever a curva e seus parâmetros, e a seleção pode ser realizada com base nestes parâmetros. Entretanto, Oliveira (1995), ao trabalhar com curva de crescimento de vacas Guzerá no Brasil, concluiu que será difícil alterar a curva de por meio de seleção praticada sobre os parâmetros da curva uma vez que a taxa de maturação e o peso assintótico apresentam correlações altas e negativas.

Segundo Fitzhugh Jr (1976), a possibilidade de alterar geneticamente a curva de crescimento depende do grau de flexibilidade genética da forma da curva, a qual, por sua vez, requer a independência entre seus parâmetros tais como de tamanho e ponto de inflexão. Ainda neste trabalho, são revisadas várias alternativas para alterar a forma da curva de crescimento, como: seleção para tamanho ou taxa de crescimento; razão entre peso e altura; escores de conformação.

Segundo Albuquerque e Fries (1998), a avaliação genética para crescimento em bovinos de corte, é realizada considerando-se os pesos ou ganhos de peso em diferentes idades como sendo características diferentes mas correlacionadas, empregando-se análises uni ou multi-características, usando-se modelos mistos. Em geral, os pesos são tomados em idades consideradas “padrão” como ao nascer, à desmama, um ano de idade e sobreano. De forma geral, ainda segundo estes autores, medindo-se e selecionando-se para alguns pontos da curva de crescimento, ou selecionando-se para os parâmetros com interpretação biológica de modelos não lineares, é possível alterar a inclinação da curva de crescimento, mas não permite alterar, significativamente, a sua forma.

2.2- Modelos de regressão não linear

Segundo Forni (2007) em rebanhos de gado de corte, o crescimento dos animais têm sido avaliado pelas medidas de peso corporal em idades padronizadas como nascimento, desmama, ano e sobreano, ou pelos ganhos de peso entre essas idades. Ainda segundo esse autor sabe-se que os pesos nessas diferentes idades apresentam correlações genéticas altas e positivas. Assim, a seleção dos reprodutores pode ser feita com base em informações obtidas no início do processo de crescimento e o aumento do peso adulto dos animais pode ser esperado e, conseqüentemente, o aumento do custo de manutenção das matrizes.

Uma forma prática e eficiente de se analisar o crescimento do animal é por meio de modelos de regressão não-lineares, como por exemplo, o modelo de Brody, que se apresenta como um modelo

adequado para descrição do crescimento de bovinos Nelore, como já mostrado no capítulo 1. Segundo Forni (2007), o modelo proposto por BRODY (1945) considera que a taxa de crescimento em cada ponto é proporcional ao crescimento que está por vir. Nesse modelo, o ponto de inflexão coincide com o nascimento, resultando em uma curva não sigmóide. Esta função foi proposta inicialmente para descrever a fase de auto-inibição do crescimento. Entretanto, é amplamente utilizada em bovinos de corte que já nascem com desenvolvimento fisiológico bastante avançado quando comparados às outras espécies, como se já tivessem superado a fase de aceleração descrita pelo parâmetro de inflexão, que é diferente de um nas demais funções. Esse modelo pode ser descrito da seguinte forma, de acordo com Jaffrézic et al. (2006):

$$y_{ij} = A_i - B_i e^{-K_i t_{ij}} + \varepsilon_{ij}, \quad i = 1, 2, \dots, N \text{ e } j = 1, 2, \dots, n,$$

O parâmetro A é o valor assintótico da função quando t (tempo, idade) tende ao infinito (sendo N o número de animais e n é o número de medidas por animal). Genericamente, este valor representa o peso adulto do animal, independente de problemas de flutuações ou, mesmo, de efeitos climáticos. O parâmetro B, ou constante de integração, não possui interpretação biológica e é utilizado para adequar o valor inicial do peso vivo, nesse caso, peso ao nascer da função utilizada. O parâmetro K é uma função entre a taxa média de crescimento e o peso adulto do animal, geralmente sendo referido como taxa de maturidade. A taxa de maturidade pode ser entendida como um indicador da velocidade com que o animal se aproxima de seu tamanho adulto. Portanto, quanto maior este valor, mais precoce é o animal em termos de crescimento.

Forni (2007) afirma que o interesse no parâmetro A advém da impossibilidade de medir o peso de todos os animais na idade adulta, pois o peso ao abate é atingido antes da maturidade. Em relação ao parâmetro K, a maturação precoce é uma característica desejável na produção de bovinos de corte, uma vez que representa diminuição no tempo que o animal necessita para atingir as características desejáveis para o abate, melhorando assim a taxa de retorno da atividade pecuária.

Dos parâmetros dos modelos de crescimento podem ser obtidas as estimativas de seus componentes de (co)variância. E a partir dessas estimativas obtêm-se os parâmetros genéticos referentes aos parâmetros dos modelos de crescimento, como as herdabilidades e as correlações. A forma adequada que se tem de obter esses componentes é por meio do uso de modelos mistos.

2.3- Modelos mistos

De acordo com Nogueira et al. (2003) o melhoramento genético animal, no Brasil, tem há anos feito uso da teoria de modelos mistos (modelos de avaliação genética), que foi desenvolvida por HENDERSON (1984), com resultados notáveis. Isso tem permitido a seleção de animais com base na predição, com alta acurácia, de seus valores genéticos, para utilização desses no melhoramento. De acordo com esses autores, a alta acurácia propiciada pelos modelos mistos se deve ao fato de levarem em conta associações genéticas existentes entre os animais, quantificadas por informações de genealogia. Isso é possível pela abordagem que considera os efeitos genéticos como tendo natureza aleatória. Os modelos mistos permitem ainda que os efeitos não-genéticos (em geral, considerados de natureza fixa) sejam removidos, de maneira a obter predições dos valores genéticos livres desses efeitos. Essas predições são denominadas “BLUP”, pela designação inglesa de “melhor predição linear não-tendenciosa”, e a terminologia “modelos mistos” vem do fato de se reunir efeitos fixos e aleatórios em um único modelo.

De modo geral, o uso de modelos mistos no melhoramento genético de bovinos, considera o vetor de observações (pesos em idades padrão) como variáveis dependentes no modelo uni ou multivariada, contendo efeitos fixos e aleatórios. Assim, com o uso desses modelos, podem-se estimar os efeitos fixos (por exemplo, efeito de rebanho, sexo, ano e etc) e prever os efeitos aleatórios relativos aos animais cujos os valores genéticos deseja-se conhecer.

Um modelo misto na sua forma geral pode ser descrito da seguinte forma (Silva et al., 2008) :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

em que:

\mathbf{y} é um vetor de observações (variáveis dependentes),

$\boldsymbol{\beta}$ o vetor de efeitos fixos a serem estimados,

\mathbf{u} o vetor efeitos aleatórios a serem preditos (valores genéticos que não são observáveis),

\mathbf{X} e \mathbf{Z} são matrizes de incidência dos efeitos fixos e aleatórios, respectivamente,

\mathbf{e} é o vetor de resíduos aleatórios.

Para o modelo adotado é usual pressupor que $\mathbf{u} \sim N(0, \mathbf{G})$ e $\mathbf{e} \sim N(0, \mathbf{P})$. No caso de um modelo unicaracterística, $\mathbf{G} = \mathbf{A}\sigma_a^2$ e $\mathbf{P} = \mathbf{I}\sigma_e^2$, em que σ_a^2 é a variância genética aditiva, σ_e^2 é a variância residual sendo estes os componentes de variância necessários para os cálculos dos parâmetros genéticos e \mathbf{A} é a matriz de parentesco genética (Nogueira, et al., 2003). Já no caso de um modelo multicaracterística (v características) $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$ e $\mathbf{P} = \mathbf{I} \otimes \mathbf{P}_0$, em que \mathbf{G}_0 e \mathbf{P}_0 serão matrizes simétricas de co(variâncias) de ordem v, genéticas e residuais, respectivamente.

Nos estudos de avaliação genética com uso de modelos mistos, considerando o estudo de curvas de crescimento (Silva, et al, 2010; Forni, 2007, Freitas, 2005; Jafrezic, et al. 2006 e etc), os componentes de variância são estimados utilizando como vetor de variáveis dependentes os parâmetros estimados nos modelos não lineares, ou seja um vetor $\boldsymbol{\phi}$, em que $\boldsymbol{\phi} = [A_1, A_2, \dots, A_N, B_1, B_2, \dots, B_N, k_1, K_2, \dots, K_N]'$, sendo i o numero de animais avaliados e $i = 1, \dots, N$.

Assim, os componentes de co(variâncias) genéticos e residuais relativos aos parâmetros do modelo que serão utilizados para estimação dos parâmetros genéticos, são obtidos considerando um modelo multicaracterística, em que as características consideradas são os parâmetros do modelo. A forma de obtenção das estimativas destes componentes pode variar de acordo com a metodologia empregada.

2.4- Estimação dos componentes de variância

Sob o ponto de vista da inferência estatística frequentista, as estimativas dos parâmetros necessários em programas de seleção são obtidas considerando duas fases distintas: uma primeira correspondendo ao ajuste individual dos modelos não-lineares, de forma a obter um conjunto de estimativas de parâmetros para cada animal; e uma segunda referente à caracterização dessas estimativas como variáveis dependentes em um modelo misto (Henderson, 1984).

De acordo com Blasco et al. (2003), a estimação em duas fases distintas não é apropriada, pois a influência de efeitos genéticos e ambientais não é considerada no processo de estimação dos parâmetros de crescimento, o que, segundo estes autores, pode caracterizar em perda de informação. Segundo esses mesmo autores, uma forma eficiente de se realizar esta análise é por meio da inferência Bayesiana. Anteriormente a esse trabalho, Varona et al. (1999) já haviam comprovado, via simulação de dados, que a estimação simultânea usando a inferência Bayesiana apresentou melhores resultados que aquela obtida em duas fases distintas.

Jaffrézic et al. (2006) relatam que a inferência Bayesiana realmente apresenta grande eficiência na estimação simultânea de parâmetros de crescimento e parâmetros genéticos, porém é muito complexa, uma vez que exige a obtenção de distribuições condicionais completas e aplicação de algoritmos MCMC, cujas convergências são de difícil constatação. Como alternativa, estes autores sugerem uma extensão do algoritmo EM com o uso de um processo estocástico, denominado algoritmo SAEM (Stochastic Approximation EM algorithm) para se obter estimativas pelo método da Máxima Verossimilhança, ou seja, sob enfoque frequentista.

Nesse mesmo estudo, a aplicação do SAEM na análise genética de curvas de crescimento proporcionou resultados estatisticamente iguais aos obtidos pela inferência Bayesiana, mas com maior praticidade, uma vez que a convergência foi rápida e não houve a necessidade de se trabalhar

algebricamente com distribuições de probabilidade para obter as distribuições condicionais completas a posteriori referentes ao método Bayesiano.

Ainda de acordo com Jaffrézic et al. (2006), o SAEM é bastante robusto para valores iniciais e tem muita rapidez na convergência das estimativas de máxima verossimilhança. Ele também fornece os intervalos de confiança para todos os parâmetros estimados e usa testes de hipóteses clássicos, além de permitir uma classificação de modelos baseados em critérios de verossimilhança como o Critério de Informação de Akaike (AIC) e o da Log-Verossimilhança.

Este último comentário é de grande importância quando se deseja comparar diferentes modelos de crescimento, uma vez que este é o objetivo de muitos trabalhos envolvendo curvas de crescimento de bovinos (Silva et al., 2001; Mazzini et al., 2005 e Silva et al., 2004), os quais consistem em propor métodos de se realizar estas comparações.

Outros autores propuseram procedimentos semelhantes para estimar simultaneamente os parâmetros do modelo de crescimento juntamente com os efeitos genéticos e ambientais, com o uso do algoritmo EM, mas em todos sempre alguma inconveniência foi encontrada como relatado por Jaffrézic et al. (2006). Segundo estes autores McCulloch (1997) propôs um algoritmo híbrido para a estimação dos componentes de variância que contempla dois algoritmos MCEM (Monte Carlo – Markov Chain algoritmo EM) e MCMLE (Monte Carlo – Markov Chain integração e maximização de verossimilhança), mas este híbrido necessita de um grande número de cadeias MCMC (problema encontrado no MCEM) e muitas iterações do algoritmo EM (problema encontrado no MCMLE), não sendo então tão eficiente quando o SAEM.

Segundo Faria et al. (2007), existem vários procedimentos de estimação dos componentes de variância já propostos para aplicação em melhoramento animal, como os de máxima verossimilhança (ML) e máxima verossimilhança restrita (REML) e os bayesianos. Sendo assim, é relevante realçar o fato que o algoritmo SAEM não constitui um novo método de estimação, mas sim um algoritmo que permite estimar simultaneamente parâmetros das curvas de crescimento e parâmetros genéticos pelo método ML com o uso do algoritmo EM.

2.5 - O algoritmo EM (esperança - maximização)

O algoritmo EM foi proposto por Dempster et al (1977) e consiste em uma técnica de estimação de parâmetros via máxima verossimilhança (ML) ou máxima verossimilhança restrita (REML) e é assim denominado por alternar entre os cálculos de valores esperados condicionais e maximização da verossimilhança.

De modo geral, o algoritmo EM pode ser aplicado nos casos onde se deseja estimar um conjunto de parâmetros que descrevem uma distribuição de probabilidades conjunta. Na utilização deste método, considera-se que o conjunto de valores observados constitui um conjunto incompleto de dados. O conjunto completo contém além dos valores observados aqueles que são não observáveis, ou seja, os considerados dados faltantes.

Segundo Junger (2006), o algoritmo consiste da idéia intuitiva de lidar com dados incompletos: (1) substituem-se os valores faltantes por valores estimados, (2) estimam-se os parâmetros, (3) reestimam-se os valores faltantes considerando que os parâmetros estimados são corretos, (4) reestimam-se os parâmetros. Este processo é repetido até que um critério de convergência seja alcançado.

Pode-se considerar θ como o vetor de parâmetros a ser estimado. Os dados faltantes serão chamados de z e os observados de y . O vetor de dados completos será então descrito por x , sendo $x=(y, z)'$.

Seja $f(x|\theta)$ a função densidade de probabilidade conjunta do vetor aleatório x , dependente do parâmetro θ e correspondente ao vetor de dados completos x . A função $g(y|\theta)$ é a densidade de probabilidade do vetor aleatório y correspondente aos dados observados y . O objetivo do algoritmo EM é encontrar o valor de θ , que maximiza $g(y|\theta)$, dados os valores observados y utilizando $f(x|\theta)$ (Junger, 2006).

O algoritmo EM é um algoritmo iterativo. Cada iteração deste algoritmo envolve 2 passos: passo E (passo da esperança) e passo M (passo de maximização). Supondo que $\theta^{(k)}$ representa o valor corrente de θ após k iterações. No passo E da iteração $(k + 1)$ é calculada a esperança condicional da função log-verossimilhança de θ dados y e $\theta^{(k)}$, que pode ser descrita da seguinte forma:

$$Q(\theta|\theta^{(k)}) = E[\log L(\theta; y, z) | y, \theta^{(k)}],$$

Em que $L(\theta; y, z)$ é a função de verossimilhança de θ .

No passo M, é determinado $\theta^{(k)}$ que maximiza a esperança condicional da log-verossimilhança. Para obter essa maximização é feita a derivada de $Q(\theta|\theta^{(k)})$ em relação a cada parâmetro do vetor θ . Assim o valor máximo é dado por:

$$\theta^k = \arg \max_{\theta} Q(\theta|\theta^k).$$

Os passos E e M se repetem alternadamente até que a diferença entre os valores da verossimilhança dos dados incompletos na k -ésima e na $(k + 1)$ -ésima iterações seja tão pequena quanto se deseja.

Segundo Delyon, Lavielle e Moulines (1999) o algoritmo EM só pode ser usado em situações onde a maximização de θ por $Q(\theta|\theta^{(k)})$ possa ser calculada de forma analítica, ou seja, a esperança condicional da função de log-verossimilhança possui uma solução analítica, o que nem sempre ocorre.

2.6-Algoritmo SAEM

Segundo Jaffrézic et al. (2006), o algoritmo EM estocástico foi introduzido por Celeux and Diebolt (1985). Uma versão da aproximação estocástica foi então proposta por Delyon et al. (1999) e aprimorada por Kuhn e Lavielle (2005,2004).

De acordo com Jaffrézic et al. (2006), a idéia geral do algoritmo SAEM é substituir a fase de Esperança do algoritmo EM (passo E), por uma aproximação estocástica. Esta substituição deverá ser feita porque o cálculo da esperança condicional da função de log-verossimilhança não possui uma forma analítica de ser calculada. De acordo com Panhard e Samson (2008), é comum essa função esperança condicional não possuir forma analítica em situações que se utilizam modelos não lineares com modelos mistos. Segundo esses autores, isso acontece devido à não linearidade da função de crescimento utilizada.

Segundo Delyon et al. (1999), a substituição da fase esperança do algoritmo EM nada mais é do que divisão deste passo em dois outros passos: Um passo de simulação (passo S) e outro que seria um procedimento de aproximação estocástica em substituição da esperança que não possui forma analítica de ser calculada. O algoritmo SAEM poderá então ser sumarizado da seguinte forma:

- 1) Simulação: geração de L cadeias aleatórias dos considerados dados faltantes. Estas cadeias serão simuladas de acordo com uma distribuição condicional diretamente ou usando um algoritmo Metropolis-Hastings. Kuhn e Lavielle (2005) afirmam que a convergência do algoritmo (SAEM) pode ser consideravelmente aumentada por um acoplamento com um procedimento MCMC como o algoritmo Metropolis-hastings. Os métodos MCMC utilizam a simulação Monte Carlo para a geração de valores a partir de distribuições de probabilidades. Fazem uso também da teoria das Cadeias de Markov, que considera um sistema de dependência entre observações subsequentes geradas na simulação. Assim são geradas M cadeias de Monte Carlo para cada uma das cadeias aleatórias dos considerados dados faltantes;
- 2) Aproximação estocástica: É a substituição da esperança condicional

$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = E[\log L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}) | \mathbf{y}, \boldsymbol{\theta}^k]$ pela seguinte aproximação estocástica:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) + \gamma_k \left(\frac{1}{L} \sum_{l=1}^L \text{Log } p(\mathbf{y}, \boldsymbol{\Phi}^{[k,l]}, \mathbf{u}^{[k,l]}; \boldsymbol{\theta}) - Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k-1)}) \right) | \mathbf{y}, \boldsymbol{\theta}^{(k-1)}$$

Os dados faltantes simulados no passo 1 são atualizados nesse passo. γ_k é uma sequência de passos positivos decrescendo para 0. De acordo com Kuhn e Lavielle (2005), para todo k em N , $\gamma_k \in [0,1]$, $\sum_{k=1}^{\infty} \gamma_k = \infty$ e $\sum_{k=1}^{\infty} \gamma_k < \infty$.

3) Maximização: consiste da maximização de $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$.

Cada parâmetro do vetor $\boldsymbol{\theta}$ deveria ser obtido pela derivação da esperança condicional da função de log - verossimilhança que quando igualado a zero forneceria seu valor máximo. Como esta esperança não tem solução analítica ela será substituída pelas aproximações estocásticas dadas no passo acima, e assim obter os valores de máximo para cada parâmetro a ser estimado.

Jaffrézic et al. (2006) afirmam que o parâmetro γ_k é um parâmetro crucial neste procedimento de estimação. Ele executa uma suavização dos valores calculados da verossimilhança de uma iteração para outra e por isso acelera consideravelmente a convergência, comparada com outros processos de estimação MCMC. Durante as K primeiras iterações não há suavização de valores, pois $\gamma_k = 1$. No entanto a partir das iterações $(K + 1)$ a suavização começa a estabilizar as estimativas e a convergir mais rapidamente para as estimativas de ML. De acordo com esses autores é recomendado tomar $\gamma_k = (k - K)^{-1}$, para $k \geq (K + 1)$. A escolha da iteração número K para começar o processo de suavização depende do número de simulações usadas no passo 1.

De acordo com Delyon et al. (1999), a convergência do algoritmo SAEM, depende da escolha do tamanho dos passos γ_k e da especificação de L usado na aproximação estocástica. Ainda segundo estes autores é inapropriado começar com pequenos valores o tamanho do passo de γ_k e grande valores para o número de simulações L por cadeia. Jaffrézic et al. (2006) afirmam que $M=5$ e $10 < L < 20$ são um bom número de cadeias a serem simuladas. No entanto $L=1$ não é suficiente para obter boa acurácia na estimação do vetor $\boldsymbol{\theta}$.

Uma vantagem da aproximação do EM estocástico é que ele permanece na estrutura clássica da verossimilhança, permitindo o uso de testes de hipóteses para os parâmetros e também permitindo comparações de modelos usando critérios clássicos como os testes de razão de verossimilhança e o critério AIC.

3 - Material e Métodos

3.1 – Dados

Os dados utilizados neste estudo foram cedidos pela ABCZ (Associação Brasileira de Criadores de Zebu). Esses dados referem-se aos pesos de 410 animais, sendo 294 machos e 116 fêmeas, nascidos de 110 touros e 360 vacas. Cada animal apresenta sete pesagens diferentes, realizadas desde o nascimento até, aproximadamente, 550 dias de idade. Em todas as análises realizadas, não houve a inserção de efeitos fixos, uma vez que tais animais, estavam sujeitos as mesmas condições de manejo e época, pois todos eram pertencentes ao mesmo grupo de

contemporâneos. É importante ressaltar que esse conjunto de dados serviu apenas para ilustrar a aplicação das metodologias empregadas, principalmente em relação ao algoritmo SAEM, que não se trata de uma técnica estatística usual na área de Melhoramento Animal. Portanto, é imprescindível explicitar que diante do baixo tamanho amostral, os parâmetros genéticos obtidos não podem ser considerados como sendo estimativas acuradas dos parâmetros genéticos populacionais referentes a rebanhos da raça Nelore.

3.2 – Modelo de Brody Modificado

O modelo de crescimento não linear utilizado para avaliar o crescimento dos animais, foi o modelo de Brody, como sugerido no capítulo 1 deste trabalho. Esse modelo é descrito da seguinte forma como sugerido por Jaffrézic et al. (2006):

$$y_{ij} = A_i - B_i e^{-K_i t_{ij}} + \varepsilon_{ij}$$

em que:

y_{ij} é o peso do indivíduo i tomado no tempo j ;

A_i , B_i , K_i são os parâmetros específicos de cada indivíduo i , sendo que A_i representa a estimativa de peso à maturidade (peso adulto); K_i a taxa de maturidade (indica a velocidade de crescimento) e o parâmetro B_i não possui interpretação biológica. Como sugerido por Jaffrézic et al. (2006), pelo fato de B_i não possuir interpretação biológica, uma reparametrização foi utilizada para este parâmetro, de tal forma que $B_i = A_i - y_{i1}$, em que y_{i1} representa o peso observado do animal i na idade 1, ou seja o peso ao nascimento do animal. Outros autores já usaram essa reparametrização, chamando esse modelo de Brody modificado, como Tedeschi, et al (2000) e Silva et al, (2002);

t_{ij} é a idade correspondente à observação y_{ij} , sendo $t_{ij} = t_{ij} / 100000$;

ε_{ij} é o erro aleatório, independentemente e normalmente distribuído, com $\varepsilon_{ij} \sim N(0, \sigma^2_e)$.

Ao assumir que os parâmetros A_i e K_i , também apresentam distribuição normal, estes podem ser melhor estudados em relação a influência de efeitos genéticos e ambientais por meio do ajuste de um modelo touro, o qual se caracteriza como um caso especial do modelo misto apresentado no item 2.3.

3.3 – Modelo Touro

O modelo touro (modelo reprodutor) prediz o efeito do pai sobre o desempenho médio de sua progênie, desconsiderando o efeito da mãe. Neste modelo os parâmetros A e K de cada indivíduo, filhos de cada touro é determinado por um efeito fixo β (média do parâmetro, já que não foram considerados grupos contemporâneos) e por um efeito genético u atribuído ao touro, além de um efeito aleatório não controlado denotado por e . Considerando essas definições o vetor Φ_i para o indivíduo i , é decomposto como segue:

$$\Phi_i = X_i \beta + Z_i u + e_i, \quad i = 1, \dots, N,$$

em que:

Φ_i é um vetor de parâmetros $2N \times 1$, $\Phi_i = [A_1, A_2, \dots, A_N, K_1, K_2, \dots, K_N]'$;

X_i é uma matriz de incidência dos efeitos fixos ($2N \times 2$);

β é um vetor de efeitos fixos (2×1);

Z_i é uma matriz de incidência dos touros ($2N \times S$), sendo S o número de touros.

u é um vetor de efeitos genéticos atribuídos aos touros ($S \times 1$), $u_i \sim N(0, A \otimes G)$, sendo A a matriz de parentesco entre os touros e G (2×2) a matriz de (co)variância genética entre os parâmetros da curva. No modelo touro os u 's preditos equivalem a metade do valor genético do indivíduo (touro);

e_i é um vetor de resíduos referentes ao modelo ($2N \times 1$), $e_i \sim N(0, \mathbf{P} \otimes \mathbf{I})$, sendo \mathbf{P} (2×2) uma matriz de covariância residual entre os parâmetros do indivíduo i e \mathbf{I} é uma matriz identidade $N \times N$.

3.4 – Obtenção dos componentes de variância

Para a obtenção dos componentes de variância provenientes do ajuste do modelo touro, os dados de bovinos Nelore foram analisados por meio das metodologias SAEM e Two-Step, com o intuito de identificar possíveis diferenças entre estas distintas técnicas estatísticas.

3.4.1 – Algoritmo SAEM

O algoritmo SAEM, foi implementado por Jafrézic et al (2006), para o estudo genético de curvas de crescimento de animais Charolais. Todos os procedimentos sugeridos e disponibilizados por esses autores foram utilizados no presente trabalho.

Esse algoritmo apresenta-se como uma ferramenta frequentista apropriada para se obter as estimativas dos parâmetros do modelo (Brody modificado) juntamente com os efeitos fixos e aleatórios e seus componentes de variância, pelo modelo touro (Jafrézic et al., 2006).

Segundo Jafrézic et al. (2006), na estrutura do algoritmo EM, para a utilização do SAEM, na análise genética de curvas de crescimento, o vetor de parâmetro θ a ser estimado será $\theta = (\beta, \mathbf{G}, \mathbf{P}, \sigma_e^2)$. Os dados faltantes serão ϕ (vetor de parâmetros do modelo não observados) e \mathbf{u} (vetor de efeitos genéticos atribuído aos touros, também não observáveis). Assim tem-se: $z = (\phi, \mathbf{u})$. O vetor de dados completos é então descrito por $\mathbf{x} = (\mathbf{y}, \mathbf{z})'$, sendo \mathbf{y} o vetor de dados observados.

Foram utilizadas na implementação deste algoritmo 700 iterações. O processo de suavização paramétrica (γ_k), começou a partir da iteração 300 ($K=300$), como recomendado por Jaffrézic et al. (2006). Cada iteração será descrita com os seguintes passos, como sumarizado no item 2.6.

- 1) Simulação: foram simuladas 10 cadeias aleatórias ($L=10$) de ϕ e \mathbf{u} . Para simulação de cada cadeia foi utilizado o algoritmo Metropolis Hasting utilizando as seguintes distribuições de probabilidade: $p(\phi | \mathbf{y}, \theta^{[k-1]})$ e $p(\mathbf{u} | \mathbf{y}, \theta^{[k-1]})$. Em cada simulação M foi igual a 5. As cadeias simuladas foram substituídas iterativamente nas aproximações estocásticas dadas no passo 2.
- 2) Aproximação estocástica: segundo Jaffrézic et al. (2006), para cada parâmetro a ser estimado β , \mathbf{G} , \mathbf{P} , e σ_e^2 tem-se as seguintes aproximações estocásticas, respectivamente:

$$S_1^k = S_1^{k-1} + \gamma_k \left[\sum_{i=1}^N \mathbf{X}_i' \mathbf{P}^{-1} (\phi_i^k - \mathbf{Z}_i \mathbf{u}^k) - S_1^{k-1} \right]$$

$$S_2^k = S_2^{k-1} + \gamma_k [\mathbf{u}^{k'} \mathbf{A}^{-1} \mathbf{u}^k - S_2^{k-1}]$$

$$S_3^k = S_3^{k-1} + \gamma_k \left[\sum_{i=1}^N (\phi_i^k - \mathbf{Z}_i \mathbf{u}^k - \mathbf{X}_i \beta^k)' (\phi_i^k - \mathbf{Z}_i \mathbf{u}^k - \mathbf{X}_i \beta^k) - S_3^{k-1} \right]$$

$$S_4^k = S_4^{k-1} + \gamma_k \left[\sum_{i,j} (y_{ij} - f(t_{ij}, \phi_i^k))^2 - S_4^{k-1} \right]$$

- 3) Maximização: maximização de $Q(\theta | \theta^{(k)})$: os parâmetros que serão estimados β , \mathbf{G} , \mathbf{P} , e σ_e^2 , serão maximizados com o uso das suas respectivas aproximações estocásticas dadas no

passo 2 (já que as esperanças condicionais não possuem forma analítica de ser resolvida). Assim, as estimativas desses parâmetros serão obtidas em cada iteração da seguinte forma de acordo com Jaffrézic et al., 2006:

$$\boldsymbol{\beta}^{(k)} = \sum_{i=1}^N (\mathbf{X}_i' \mathbf{P}^{-1} \mathbf{X}_i) S_1^k$$

$$\mathbf{G}^{(k)} = \frac{S_2^k}{N_a}$$

$$\mathbf{P}^{(k)} = \frac{S_3^k}{N}$$

$$\sigma_e^{2(k)} = \frac{S_3^k}{N_{\text{tot}}}$$

em que: N é o número de indivíduos com observações, N_{tot} é o número total de observações ($N_{\text{tot}} = \sum_{i=1}^N n_i$) e N_a é o número de animais na matriz de parentescos.

Após a obtenção dos valores que maximizam a função de máxima verossimilhança neste último passo, estes são atualizados no passo 1, dando origem ao processo iterativo. Tal processo é então repetido até se obter a convergência. Em razão desta atualização constante, é dito que temos um processo de estimação conjunta (parâmetros do modelo, efeitos fixos e aleatórios e componentes de variância são estimados conjuntamente), caracterizado pela correção simultânea, isto é, a cada iteração, de um efeito sobre os demais. Idéia análoga, sob o ponto de vista da metodologia Hierárquica Bayesiana, é apresentada em Varona et al. (1999), Blasco et al. (2003) e Silva et al. (2010).

3.4.2 - Two Step

O método frequentista Two Step (dois passos) é o mais utilizado para obtenção dos componentes de variância. Vários autores utilizaram esse método em seus estudos com bovinos como Silva (2009), Silva et al. (2002), Garnero et al. (2005), Santoro et al. (2005), Talhari et al, (2003).

O método consiste em duas etapas independentes para obtenção dos componentes da variância. A primeira etapa consiste na obtenção dos parâmetros do modelo de crescimento (parâmetros A e K , do modelo de Brody modificado). Na segunda etapa as estimativas obtidas para os parâmetros do modelo da etapa 1 são utilizadas como variáveis dependentes em um modelo touro contendo os efeitos fixos e aleatórios. A partir deste modelo touro pode-se obter os componentes de variâncias necessários aos cálculos das herdabilidades e correlações.

A grande diferença dessa metodologia para a metodologia descrita anteriormente (algoritmo SAEM) é a independência no processo de obtenção dos parâmetros do modelo e dos componentes de variância.

3.5 - Valores genéticos e curvas genéticas de touros

Como já especificado neste trabalho, sempre que se utiliza o modelo misto, obtemos as predições dos valores genéticos (\mathbf{u}) para os animais avaliados. No caso do modelo touro, que foi utilizado no presente estudo, o valor obtido pelos \mathbf{u} 's corresponde, na verdade, a metade do valor genético do reprodutor.

No caso de análise genética dos parâmetros A e K dos modelos de crescimento, a qual é geralmente considerada como uma análise multicaracterística, neste caso bivariada, são obtidos valores de \mathbf{u} 's relativos ao parâmetro A , cujo vetor que contempla os valores de todos os touros é \mathbf{Ua} ;

e relativos ao parâmetro K , cujo vetor correspondente é U_k . Para testar as metodologias descritas em 3.4, os valores de U_a e U_k obtidos por ambas as metodologias foram comparados mediante correlação de Spearman, cujos coeficientes estimados representam o grau de similaridade entre as mesmas no que diz respeito à predição de tais valores.

De acordo com Albuquerque e Fries (1998), a avaliação genética dos animais baseada apenas nas características representadas pelas estimativas dos parâmetros do modelo de crescimento não é o mais usual e nem o mais recomendado para fins de seleção. Assim, o conhecimento do comportamento da trajetória da curva de crescimento genética do animal durante todo o período de tempo considerado é bem mais interessante, pois permite predizer os valores genéticos dos animais em qualquer ponto da curva, inclusive nas idades padrão (nascimento, desmama, ano e sobreano), cujas pesagens são geralmente usadas na maioria dos programas de seleção para gado de corte.

De forma geral, a metodologia de Regressão Aleatória permite a obtenção da referida curva genética, mas esta metodologia utiliza-se apenas de modelos lineares, como os polinômios de Legendre, cujos parâmetros não apresentam interpretação biológica como aqueles dos modelos de crescimento não lineares. Dessa forma, uma alternativa viável seria explorar a interpretação dos parâmetros dos modelos de regressão não linear sob o ponto de vista biológico, e simultaneamente ter o conhecimento do comportamento da trajetória da curva de crescimento genética do animal.

A obtenção da curva de crescimento genética dos animais pode ser obtida conforme metodologia apresentada por Pong-Wong e Hadjipavlou (2009), com algumas adaptações.

3.5.1 - Metodologia adaptada de Pong-Wong e Hadjipavlou (2009)

Pong-Wong e Hadjipavlou (2009), utilizaram o modelo de regressão não linear de Gompertz para fazer a avaliação genômica em ovelhas. No presente trabalho, foi feita uma adaptação na metodologia proposta por esses autores, uma vez que não foram consideradas informações moleculares, para a obtenção de curvas genéticas dos touros avaliados.

Assim, mediante a adaptação da metodologia Pong-Wong e Hadjipavlou (2009), para a obtenção das curvas genéticas de touros, essa pode ser descrita da seguinte forma:

- 1) Obtem-se os valores de U_a e U_k preditos pelo modelo touro ;
- 2) Estes valores foram somados aos valores médios estimados de seus respectivos parâmetros de forma a obter um novo vetor U_a^* e U_k^* da seguinte forma:

$$U_a^* = \mu_a + U_a \quad e \quad U_k^* = \mu_k + U_k ;$$

- 3) Os vetores U_a^* e U_k^* foram substituídos no Modelo de Brody modificado (item 3.2), juntamente com valores de idade de forma a obter uma curva de crescimento genética para cada touro da seguinte forma:

$$y_{ij}^* = U_a^* - ((U_a^* - PN_i)e^{-U_k^* \left(\frac{t_{ij}}{100000}\right)},$$

em que:

y_{ij}^* é o vetor de valores genéticos acrescidos da média para os pesos do touro i em cada idade j ;

U_a^* e U_k^* são os vetores de valores genéticos acrescidos da média para as estimativas dos parâmetros A e K ;

PN é a média do peso ao nascimento para os filhos touro i ;

t_{ij} são as pesagens simuladas j relacionadas ao touro i , sendo que t variou de 0 a 550 dias de idade, com um intervalo de 10 em 10 dias.

Além da estimação da curva genética, foi estimada também uma curva média para os touros, considerando o mesmo modelo apresentado no passo 3. A diferença é que nesta curva média foram

considerados apenas os valores estimados pelos efeitos fixos (μ_a e μ_k), ou seja, não foram utilizados os valores genéticos. A curva média é descrita então da seguinte forma:

$$y_{ij}^{**} = \mu_a - ((\mu_a - PN_i)e^{-\mu_k(\frac{t_{ij}}{100000})}).$$

A partir da diferença entre os valores gerados pela curva média (y_{ij}^{**}) e a curva genética estimada para cada touro (y_{ij}^*), tem-se os valores genéticos preditos para cada touro em qualquer idade de interesse (do nascimento aos 550 dias). Por meio desta técnica, foram obtidos os valores genéticos destes touros em idades padrão: nascimento (0 dias), desmama (210 dias), ano (360 dias) e sobreano (540 dias). Os valores genéticos obtidos para as essas idades foram então correlacionadas entre as diferentes metodologias testadas a fim de verificar semelhança das mesmas na predição de tais valores.

3.5.2 - Análises unicaracterísticas para pesos observados em idades padrão

Além dos valores genéticos obtidos a partir da metodologia descrita, também foram obtidos valores genéticos por meio do ajuste do modelo touro em análises unicaracterísticas para os pesos observados nas idades próximas as idades padrão. Tais valores genéticos foram correlacionados por meio do coeficiente de Spearman com os valores genéticos obtidos para as idades padrão pela metodologia adaptada de Pong-Wong e Hadjipavlou (2009) ao se considerarem as metodologias SAEM e Two Step .

3.6 – Análises realizadas

Os métodos SAEM e Two Step, foram aplicados aos dados de bovinos descritos neste trabalho, com o intuito de verificar possíveis diferenças entre as metodologias, tendo em vista a estimação dos componentes de variância, parâmetros genéticos, efeitos fixos e valores genéticos dos animais. A forma de aplicação destes métodos se deu de quatro maneiras distintas, conforme descrições a seguir:

1ª análise) SAEM1- Os dados dos 410 animais Nelore, foram analisados a partir do algoritmo SAEM. Como já relatado, pelo fato desse algoritmo realizar uma análise conjunta, não há problemas de convergência dos parâmetros do modelo e todos os animais foram incluídos em ambos os processos de estimação (parâmetros do modelo de crescimento e parâmetros genéticos). Assim, essa análise foi realizada com 410 animais e 110 pais (já que se tratava de um modelo touro)

2ª análise) Two- Step 1 (TS1)- Dos 410 animais utilizados na primeira etapa de ajuste deste método, apenas 326 obtiveram convergência para os parâmetros do modelo. Assim, para a segunda etapa do processo, somente foram utilizados as estimativas de parâmetros de 326 animais consequentemente reduzindo o número de pais para avaliação no modelo touro para 89.

3ª análise) SAEM2- Mediante a diminuição no número amostral no TS1, e para fins de comparação das metodologias utilizando o mesmo número de animais, o algoritmo SAEM foi utilizado para analisar apenas os dados dos mesmos 326 animais que convergiram no segundo método de análise (TS1). Assim, o objetivo deste segundo ajuste via SAEM é obter resultados, independentes da diferença do número de animais envolvidos em relação a análise TS1. O número de pais referentes ao conjunto de dados 326 animais foi de 89 pais.

4ª análise) Two- Step 2 (TS2)- Ainda sob o enfoque de comparar as metodologias utilizando os mesmos números de animais, e com o intuito também de verificar se a forma de estimação conjunta dos parâmetros do modelo influencia na estimação dos componentes de variância, uma segunda alternativa para o Two-step foi avaliada. Nesta análise, os parâmetros do modelo (A e K) estimados pelo primeiro método (SAEM 1), foram utilizados como as variáveis dependentes no modelo misto (modelo touro), de forma que, a partir destes parâmetros, já estimados conjuntamente pelo SAEM,

fosse possível obter os componentes de variância, de forma usual (método two step). Assim, essa análise foi realizada considerando as estimativas dos parâmetros do modelo de crescimento dos 410 animais na análise SAEM1, sendo que os componentes de variância estimados no segundo estágio caracterizaram-se por serem independentes dos valores assumidos para A e K no primeiro estágio.

3.7 – Aspectos Computacionais

Para a realização das análises com o algoritmo SAEM, foram utilizadas as rotinas computacionais disponibilizadas pelo segundo autor (Cristian Meza) do artigo de Jaffrézic et al., (2006), Essas rotinas estão na homepage deste autor: <http://staff.deuv.cl/meza/softBIS.htm>, e encontram-se implementadas no programa computacional Matlab (2007). É importante ressaltar, que Cristian Meza, disponibiliza na sua homepage todo o processo de simulação usado no artigo de Jaffrézic et al. (2006), para verificação da validade do algoritmo SAEM. Pelas comprovações já mostradas por esse autor em seu trabalho e pela disponibilização do arquivo de simulação, não foi julgado necessário fazer um novo estudo de simulação no presente trabalho.

Para a realização das análises pelo método Two-Step, dois programas computacionais foram utilizados, um para a análise da primeira fase e outro para a da segunda fase. Na primeira fase foi utilizado o programa computacional SAS® (SAS, 2003), que por meio do PROC MODEL ajustou o modelo de regressão não linear de Brody modificado aos dados de crescimento dos bovinos Nelore. Para a segunda fase deste método, foi utilizado o programa computacional WOMBAT (Meyer, 2006), o qual é um programa computacional adequado para a análise genética envolvendo modelos mistos uni e multivariados pelo método de REML. Ao se considerar o modelo touro bivariado, este último software já fornece diretamente as estimativas dos componentes de variância genética aditiva e residual para facilitar o cálculo de herdabilidades e correlações genéticas, conforme exemplo 2B disponível na página <http://didgeridoo.une.edu.au/km/wmbdownloads.php>.

4-Resultados e Discussão

4.1-Componentes de variância e parâmetros genéticos

Na tabela 1 estão apresentadas as estimativas dos componentes de co(variância) genética e residual obtidas nas quatro análises realizadas.

Tabela 1- Matrizes de co(variância) dos efeitos genéticos e residuais e médias (μ) estimadas dos parâmetros A e K para os quatro análises realizadas

	Genético	Residual	μ_A	μ_K
SAEM1	$\begin{bmatrix} 7689,9 & -2423,9 \\ -2423,9 & 1495,2 \end{bmatrix}$	$\begin{bmatrix} 7722,9 & -526,48 \\ -526,48 & 110,67 \end{bmatrix}$	503,72	184,99
TS1	$\begin{bmatrix} 8639,3 & -8542,0 \\ -8542,0 & 8904,5 \end{bmatrix}$	$\begin{bmatrix} 21310,4 & -10362,9 \\ -10362,9 & 12657,9 \end{bmatrix}$	475,4	253,05
SAEM2	$\begin{bmatrix} 4152,6 & -409,46 \\ -409,46 & 133,62 \end{bmatrix}$	$\begin{bmatrix} 6299,5 & -1885 \\ -1885 & 1051,4 \end{bmatrix}$	431,38	232,1
TS2	$\begin{bmatrix} 7962,5 & -74,76 \\ -74,76 & 492,32 \end{bmatrix}$	$\begin{bmatrix} 11160,8 & -2543,87 \\ -2543,87 & 1123,9 \end{bmatrix}$	498,97	191,15

A partir dos resultados apresentados na tabela 1, pode-se perceber que em relação aos efeitos fixos estimados, não houve grande diferença entre as análises realizadas. Esses valores estimados representaram na verdade, a média de estimativa de cada parâmetro, pelo fato de não se ter

considerado nenhum efeito fixo, como por exemplo, a formação de grupos contemporâneos. Os valores encontrados ficaram condizentes com Silva et al, (2002), que também utilizaram a função de Brody modificada com bovinos Nelore.

Em relação aos componentes de variância, apresentados na tabela 1, pode-se perceber que o SAEM1 apresentou uma grande diferença em relação ao TS1, principalmente em relação aos componentes residuais. A princípio, pode se pensar que esta diferença se deve ao fato do SAEM1 ter envolvido um maior número de indivíduos (410 animais e 110 touros) que o método TS1 (326 animais e 89 touros) devido a problemas de convergência na primeira etapa deste segundo método. Contudo, pelo fato do TS1 ser executado em duas etapas distintas, os erros provenientes do ajuste do modelo de crescimento na primeira etapa não são considerados na estimação dos componentes de variância na segunda etapa, o que pode também estar elevando essa variação residual. Como no SAEM1 a estimação é simultânea, há uma correção nas estimativas dos componentes de variância, diminuindo a variação residual.

Em analogia ao parágrafo anterior, ao se comparar as metodologias TS1 e SAEM2, as quais utilizaram os mesmos números de animais, verifica-se que a referida diferença permanece, indicando então que a correção simultânea é a responsável pelas diferentes estimativas obtidas. Percebe-se ainda, que o algoritmo SAEM, indiferente do número de animais (SAEM1 e SAEM2), gera estimativas análogas para os componentes de variância, mostrando que a correção simultânea assegura a baixa instabilidade na estimação dos componentes de variância.

Também é possível notar na tabela 1 que quando comparamos o SAEM1 com o TS2, as estimativas das matrizes de covariância se apresentam bem próximas. Isto significa que ao utilizar as estimativas de A e K do SAEM1 para se estimar os componentes de variância pelo método TS (TS2), as estimativas dos componentes de variância obtidas se aproximaram. Esse fato elucida que a diferença na estimação dos componentes de variância é realmente proveniente do erro de ajuste da primeira etapa.

Para concluir as discussões da tabela 1, de forma geral, percebe-se que as análises SAEM1 e SAEM2 são concordantes, pois ambas constituem a mesma metodologia cujos resultados mostram-se homogêneos ao se considerar diferentes números de animais. Concomitantemente, as análises SAEM1 e TS2 também são concordantes, uma vez que os fenótipos considerados em TS2 foram provenientes da correção simultânea executada pelo SAEM1.

No entanto é importante também considerar em relação às discussões da tabela 1 que um outro fator que pode estar elevando as diferenças entre os componentes de variância estimados, é o fato do SAEM e TS considerarem metodologias diferentes para estimação desses componentes. O SAEM utiliza o método de ML e o TS o de REML. Por esse motivo as estimativas do SAEM podem estar sendo subestimadas, pois esse já um comportamento esperado quando se utiliza a ML. Meza et al (2007), comprovam esse fato em um estudo de simulação de dados, onde estes autores compararam estimativas de componentes de (co)variância dos parâmetros da curva de crescimento de Gompertz pelos métodos ML e REML. Esses autores obtiveram, conforme o esperado, menores valores quando utilizada a ML principalmente para os componentes genéticos em relação aos residuais. Assim, pode-se concluir que as diferenças observadas podem na verdade ser menores que as observadas na tabela 1 pelo viés de estimativa quando se considera as estimativas de ML.

Na tabela 2 são apresentadas as estimativas de herdabilidades para os parâmetros A e K e correlação genética entre estes eles obtidas pelos quatro análises consideradas.

Tabela 2- Valores calculados de herdabilidades (h^2_A e h^2_K) e correlação genética ($r_g(A, K)$) para os parâmetros A e K

	h^2_A	h^2_K	$r_g(A, K)$
SAEM1	0,49	0,93	-0,71
TS1	0,28	0,41	-0,97
SAEM2	0,39	0,11	-0,54
TS2	0,41	0,30	-0,03

Pela tabela 2, tendo em vista as estimativas de herdabilidade obtidas para o parâmetro A, pode-se dizer que os valores são concordantes com a conclusão da tabela 1 relativa as semelhanças entre as análises SAEM1 e SAEM2 e entre as análises SAEM1 e TS2. Para o parâmetro K esta semelhanças não foram tão evidentes principalmente pela alta estimativa de herdabilidade pelo método SAEM1.

De forma geral, os resultados da tabela 2 são em parte concordantes com os resultados relatados na literatura. Silva et al. (2002) em suas revisões sobre valores de herdabilidade de parâmetros de modelos de crescimento aplicados a bovinos de corte, relatam valores de herdabilidade de A variando entre 0,04 a 0,46 e de K variando entre 0,08 a 0,41. Garnero et al. (2005) estudaram o crescimento de fêmeas Nelore pelos modelos de Brody, Gompertz, Logístico e Von Bertalanffy e concluíram que os coeficientes de herdabilidade foram de elevada magnitude para o parâmetro A (0,39-0,42) e também para o parâmetro K (0,39-0,44), sugerindo que estes parâmetros podem ser utilizados no estabelecimento de programas de seleção. Forni (2006) estudando o crescimento de bovinos Nelore pelo Modelo de Von Bertalanffy, encontrou valores de herdabilidade para A de alta magnitude (0,52) e para o parâmetro K de baixa magnitude (0,06).

Os valores de correlação genética com exceção do método TS2 (tabela 2), também condizem com a literatura (Garnero et al., 2005), sendo altos e negativos. Em relação ao valor encontrado para esse parâmetro na análise TS2, Santoro et al. (2005) e Silva et al (2010) que encontraram correlações genéticas positivas entre os parâmetros A e K para animais Nelore, afirmam que este é um comportamento de difícil explicação e, possivelmente, se devem a problemas de amostragem.

4.2 – Valores genéticos para os parâmetros do modelo

As figuras 1 e 2 apresentam, respectivamente, os gráficos de dispersão dos valores genéticos (U_a e U_k) obtidos pelas análises realizadas para os parâmetros A e K. Em cada um dos gráficos foi colocado o valor do coeficiente de correlação de Spearman, juntamente com sua significância.

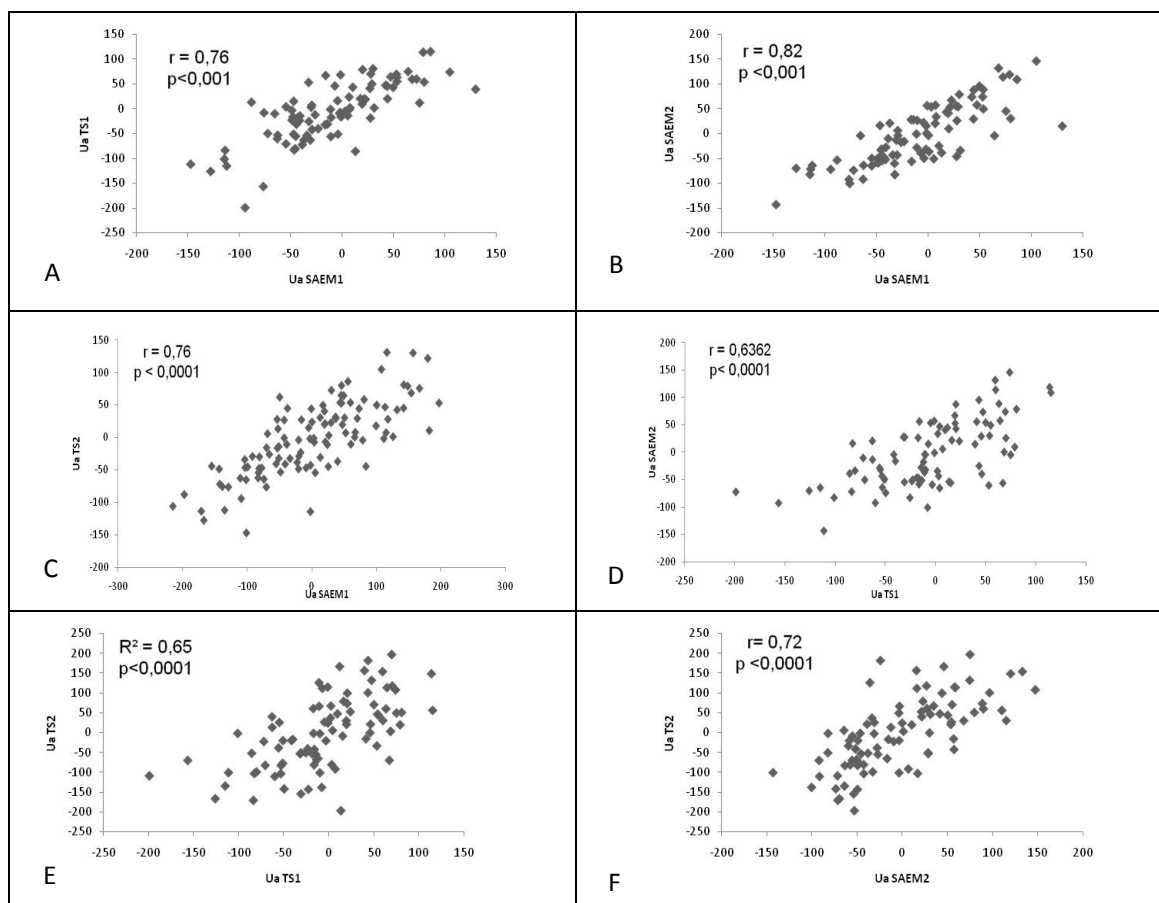


Figura 1- Gráficos de dispersão e valores do coeficiente de Spearman entre os valores genéticos preditos para o parâmetro A (U_a) nas quatro análises realizadas.

De forma geral, pode-se perceber pela figura 1 que embora tenham sido observadas diferenças nos componentes de variância (tabela 1) e nos parâmetros genéticos (tabela 2), os valores genéticos para o parâmetro A tenderam a ser semelhantes de forma global no rank de seus valores, uma vez que as correlações de Spearman variaram entre 0,63 e 0,82 e foram todas significativas. Percebe-se ainda, que aquelas análises que tiveram maiores semelhanças na estimação dos componentes de variância (tabela 1) foram as que apresentaram maiores valores de correlação entre os valores genéticos para A (figura 1B e 1C), confirmando então a concordância do SAEM em seu processo de estimação e a baixa precisão do TS na primeira etapa do método.

Para os valores genéticos do parâmetro K, pode-se observar pela figura 2 que esses valores também foram significativos, embora de menor magnitude, variando entre 0,51 e 0,87.

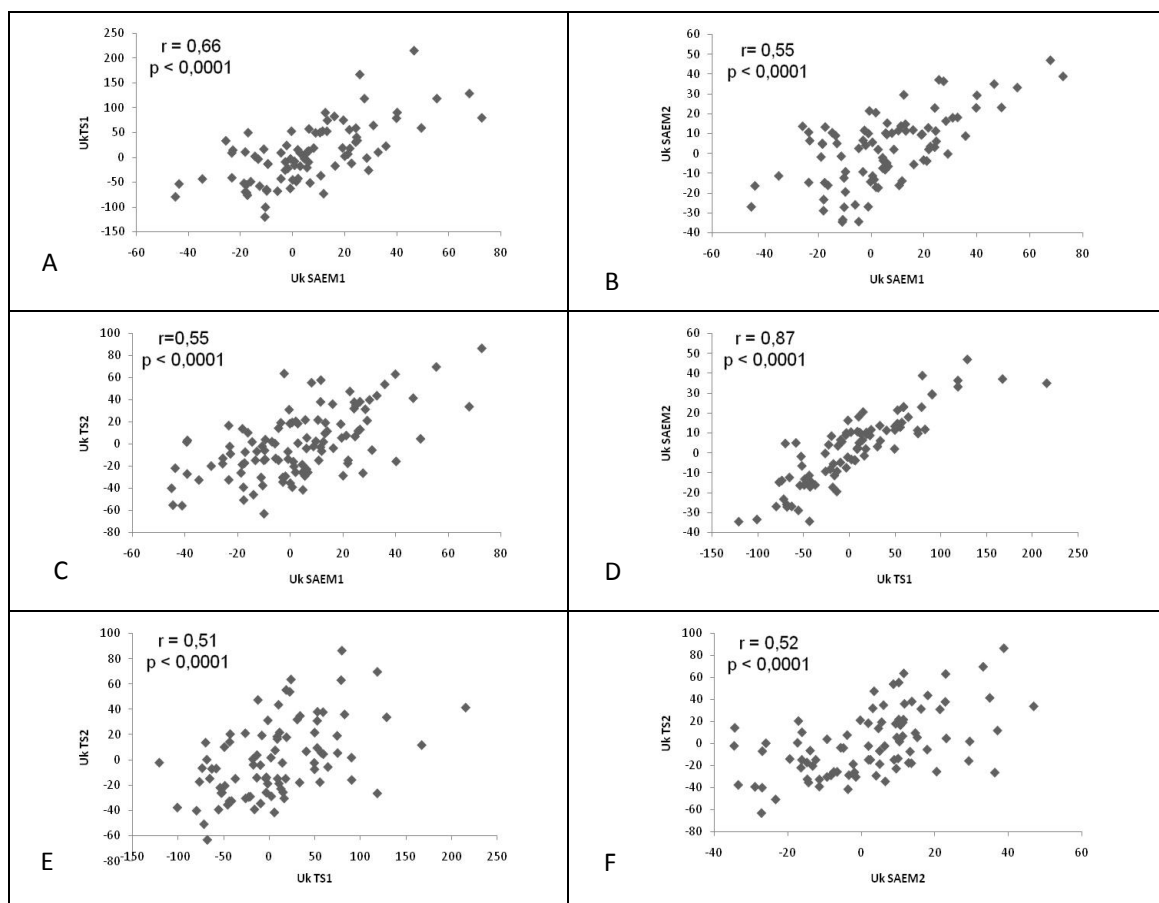


FIGURA 2- Gráficos de dispersão e valores do coeficiente de Spearman Spearman entre os valores genéticos preditos para o parâmetro K (Uk) nas quatro análises realizadas.

Na tabela 3 são apresentados os melhores e os piores touros em relação aos valores genéticos para as quatro análises realizadas, e na tabela 4 quantos desses touros foram comuns em cada comparação entre as análises tomadas duas a duas, ou seja, evidencia a concordância de classificação dos touros entre as análises consideradas.

Tabela 3 – Ordenação dos dez touros superiores e inferiores em relação aos valores genéticos preditos para os parâmetros A e K nas 4 análises realizadas

	A				K			
	SAEM1	TS1	SAEM2	TS2	SAEM1	TS1	SAEM2	TS2
10 +	108	101	43	13	28	17	25	28
	52	49	71	4	25	86	28	96
	66	54	49	66	96	25	86	90
	43	85	32	44	34	96	26	6
	101	80	101	52	17	26	17	100
	2	43	62	71	87	87	96	53
	81	19	11	49	6	22	22	104
	49	13	65	2	104	39	87	56
	44	91	54	63	20	28	34	20
	32	3	13	15	61	6	6	17
10-	99	39	25	99	67	91	64	49
	86	28	6	96	33	13	19	40
	97	6	17	42	52	11	11	55
	17	22	45	45	46	3	43	80
	100	87	87	53	37	19	80	23
	96	26	67	90	8	54	13	63
	6	96	75	25	66	85	55	19
	87	25	86	6	81	80	49	2
	25	86	42	97	2	49	32	66
	26	17	26	100	80	101	101	13

Tabela 4- Concordância dos métodos de análise considerando comparações 2 a 2 na ordenação dos animais

Comparações	10+ A	10+ K	10- A	10- K	Média
SAEM1 x TS1	3	6	7	1	4,25
SAEM1 x SAEM2	4	7	6	1	4,50
SAEM1 x TS2	5	6	6	3	5,00
TS1 x SAEM2	5	9	6	6	6,50
TS1 x TS2	2	4	3	4	3,25
SAEM2 x TS2	3	4	4	5	4,00
média	3,66	6,00	5,34	3,34	4,58

Pode-se perceber pelas tabelas 3 e 4, que as análises que mais concordaram em relação à classificação dos touros foram aquelas que usaram o mesmo número de animais, SAEM1 x TS2 e TS1 x SAEM2, as quais apresentaram média, respectivamente, de 50 e 65% dos animais coincidentes. Até certo ponto, este é realmente o resultado esperado, pois a probabilidade de se observar animais coincidentes em conjuntos de dados semelhantes é maior que aquela proveniente de conjuntos de dados diferentes.

Dentre as comparações que utilizavam animais diferentes, pode-se perceber que a comparação entre as duas análises com o algoritmo SAEM (SAEM1 e SAEM2) foram aquelas que sugeriram o maior número de indivíduos com valores genéticos parecidos, sugerindo então a maior estabilidade deste método na indicação dos animais para seleção. Em termos práticos, mesmo com conjuntos de dados diferentes, os selecionados na SAEM1 tenderam a também serem selecionados ao se usar o SAEM2.

Ainda de acordo com as tabelas 3 e 4, pode-se perceber que os indivíduos com maiores valores genéticos para K e menores para A foram os que mais concordaram em média (valores 6,00 e 5,34 na última linha tabela 4) entre as diferentes análises, quando comparadas duas a duas. Para fins de seleção, esse fato é importante, pois indivíduos com maiores valores de **U_k** são indicados para

aumentar geneticamente a precocidade do rebanho a ser melhorado. Assim, entre as 4 diferentes análises consideradas, aquelas que mais concordaram para esta característica (maiores valores de Uk) foi TS1 x SAEM2, uma vez que foram observadas 9 coincidências. Esta alta coincidência pode ser decorrente do fato destas metodologias utilizarem o mesmo número de touros (89), levando as mesmas considerações enfatizadas nos dois parágrafos anteriores.

Para confirmar os altos valores de correlação genética entre A e K já apresentados e discutidos anteriormente (tabela 1) para SAEM1, TS1 e SAEM2, o número de coincidências para cada uma destas análises entre menores valores genéticos para A e maiores para K (tabela 3) foram computados, sendo que o TS1 indicou todos os animais como coincidentes. Tal análise foi seguida pelas análises SAEM1 e o SAEM2 que, respectivamente, apresentaram 6 e 5 coincidências. Ao se pensar em propostas para seleção, é interessante realçar que em muitos rebanhos não é de interesse aumentar, por meio de cruzamentos direcionados, o peso adulto dos animais, pois isto implica em maior gasto com a manutenção, e tal afirmação diz respeito principalmente a vacas dispostas em um sistema de bovinocultura de corte. Além disso, sabe-se também que no sistema em questão, a precocidade dos animais é uma das principais características requeridas. Portanto os touros que apresentam altos valores genéticos para K são os mais indicados sob este ponto de vista.

4.3 - Curvas genéticas e valores genéticos em idades padrão

A partir da metodologia adaptada de Pong-Wong e Hadjipavlou (2009), foram obtidas para cada análise realizada, as curvas genéticas da trajetória de crescimento de cada touro e a curva média de crescimento. Essas curvas podem ser visualizadas nas figuras 3 a 6.

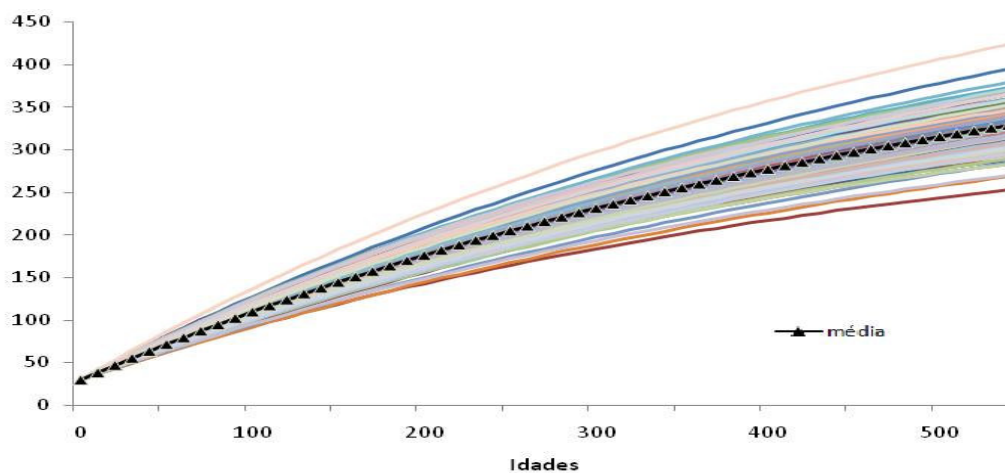


Figura 3 - Curvas genéticas individuais para touros e curva média obtidas pela análise SAEM1

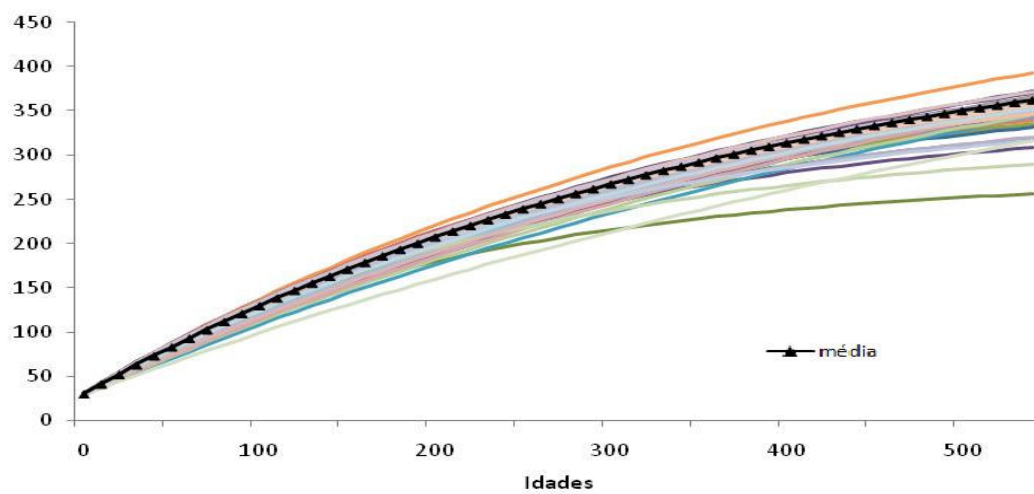


Figura 4 - Curvas genéticas individuais para touros e curva média obtidas pela análise TS1.

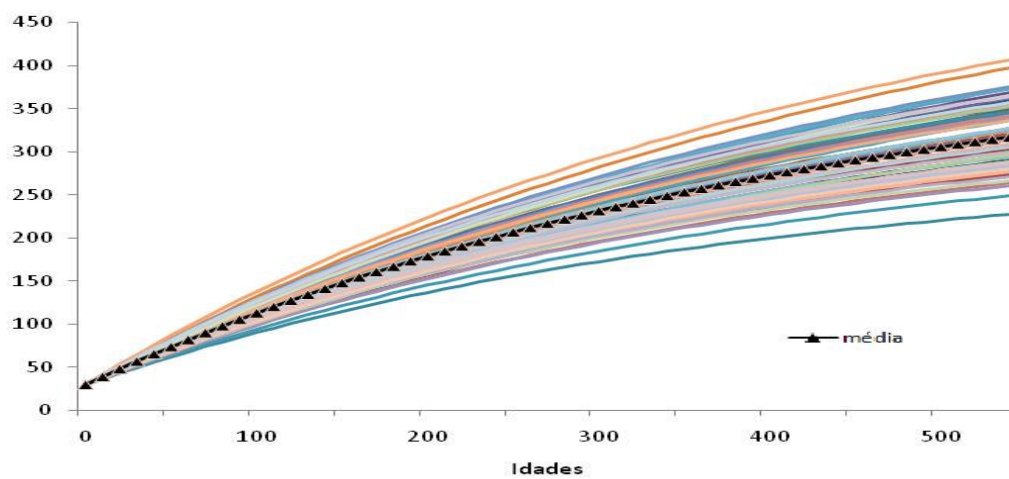


Figura 5 - Curvas genéticas individuais para touros e curva média obtidas pela análise SAEM2

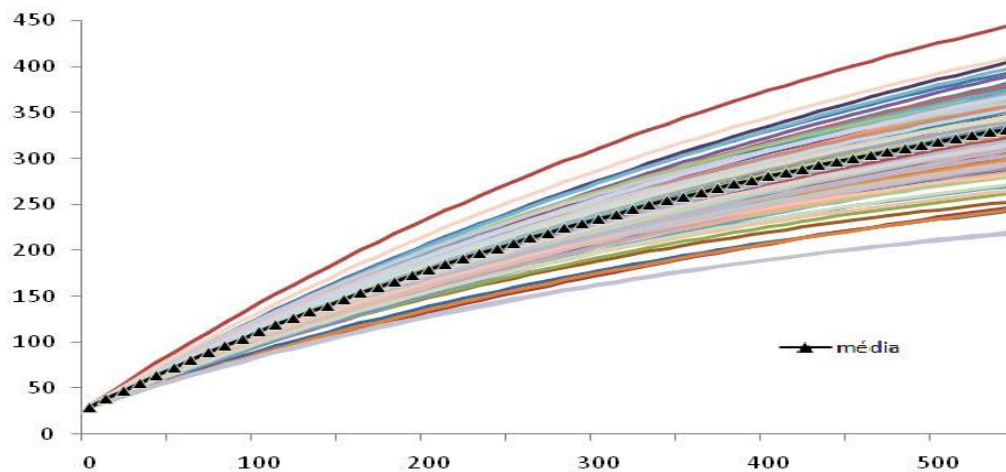


Figura 6- Curvas genéticas individuais para touros e curva média obtidas pela análise TS2

Pode-se perceber pelas figuras 4 e 6, que de forma geral, as curvas genéticas obtidas pelas análises TS1 e TS2 apresentam diferenças em relação a variação das curvas em torno da curva média, sendo que esta diferença marcante quanto a variação não é observada nos resultados obtidos pelas análises SAEM1 (figura 3) e SAEM2 (figura 5), corroborando as discussões das tabelas 1 e 8, nas quais verifica-se a maior estabilidade do método SAEM.

Conforme relatado no item 3.5, foram obtidos para cada touro os valores genéticos em idades padrão (nascimento, desmama, ano e sobreano), sendo que tais valores representam as diferenças entre os valores genéticos da curva genética individual e os valores médios. Esses valores foram correlacionados através do coeficiente de Spearman, para todas as análises realizadas, como mostra a tabela 4. Pode-se observar por essa tabela, que para todas as idades avaliadas, as maiores correlações foram encontradas entre SAEM1 x SAEM2 (sempre iguais a 95%) seguidas por SAEM1 x TS2 (na faixa de 80%). Tais resultados também confirmam mais uma vez a semelhança, ou a já referida estabilidade, das análises baseadas no algoritmo SAEM.

Observa-se ainda (tabela 4) que os coeficientes de correlação tendem a se tornar maiores a medida que as idades padrão aumentam. Dessa forma pode-se considerar que a seleção feita em idades mais avançadas (a partir de um ano), tende a ser menos influenciada pelo método em questão. Além disso, pode-se relacionar tal resultado com as curvas mostradas nas figuras 3 a 6, nos quais nota-se que a variância dos valores genéticos em cada tempo aumenta com o avançar da idade.

Tabela 5- Valores do coeficiente de correlação de Spearman entre os valores genéticos das idades padrão, nas quatro análises realizadas

Peso ao nascimento (PN)				
	UPN-SAEM1	UPN -TS1	UPN -SAEM2	UPN -TS2
UPN -SAEM1	1	-0,007 ^{NS}	0,95**	0,80**
UPN -TS1		1	-0,18 ^{NS}	-0,28*
UPN -SAEM2			1	0,85**
UPN -TS2				1
Peso a desmama (PD)				
	UPD-SAEM1	UPD -TS1	UPD -SAEM2	UPD -TS2
UPD -SAEM1	1	-0,003 ^{NS}	0,95**	0,81**
UPD -TS1		1	-0,09 ^{NS}	-0,19*
UPD -SAEM2			1	0,85**
UPD -TS2				1
Peso ao Ano (PA)				
	UPA-SAEM1	UPA -TS1	UPA -SAEM2	UPA -TS2
UPA-SAEM1	1	0,13 ^{NS}	0,95**	0,83**
UPA -TS1		1	0,12 ^{NS}	0,11 ^{NS}
UPA -SAEM2			1	0,84**
UPA -TS2				1
Peso ao Sobreano (PS)				
	Ups-SAEM1	UPS -TS1	UPS -SAEM2	UPS -TS2
UPS-SAEM1	1	0,42**	0,95**	0,83**
UPS -TS1		1	0,43*	0,50**
UPS -SAEM2			1	0,83**
UPS -TS2				1

** p<0,0001, *p<0,001, NS= não significativo

Para verificar a semelhança dos valores genéticos obtidos pelas quatro análises realizadas com os valores obtidos a partir de análises unicaracterísticas para os touros em idades padrão (item 3.5.2), foram calculadas as correlações de Spearman entre esses valores os quais estão na tabela 5.

Tabela 6- Valores do coeficiente de correlação de Spearman entre os valores genéticos das idades padrão dos touros pela análise unicaracterística e pelas metodologias testadas

89 pais							
PN		PD		PA		PS	
TS1	SAEM2	TS1	SAEM2	TS1	SAEM2	TS1	SAEM2
-0,01 ^{NS}	0,48*	0,02 ^{NS}	0,63**	0,15 ^{NS}	0,93**	0,46**	0,95**
110 pais							
PN		PD		PA		PS	
SAEM1	TS2	SAEM1	TS2	SAEM1	TS2	SAEM1	TS2
0,43**	0,45**	0,68**	0,52**	0,92**	0,82**	0,93**	0,84**

** p<0,0001, *p<0,001, NS= não significativo

Pelos resultados apresentados na tabela 5, pode-se verificar que o algoritmo SAEM indiferentemente do número de touros considerado, foi o que apresentou maiores valores para a correlação de Spearman com os valores genéticos obtidos pelas análises unicaracterísticas para cada idade padrão. A análise TS2 também apresentou valores de magnitude média a alta para este coeficiente, indicando que a grande diferença das metodologias esta no primeiro estágio de análise do TS. Pode-se observar ainda por esta tabela, que os valores de correlação, são menores no nascimento e na desmama, sendo maiores no ano e sobreano.

Além disso, de forma geral, pode-se relatar que independentemente da metodologia usada, SAEM ou Two Step, o método proposto por Pong Wong e Hadjipavlou (2009) apresenta como uma

importante ferramenta para o melhoramento genético moderno, pois possibilita obter estimativas de valores genéticos para qualquer idade de interesse.

5-Conclusões

De modo geral, resumindo os objetivos propostos para o presente trabalho, pode-se concluir que o algoritmo SAEM se mostrou adequado para a proposta de avaliação genética de bovinos, pois apresentou estimativas mais estáveis quando comparadas ao TS, além de ter predito valores genéticos para idades padrão correlacionados significativamente com aqueles obtidos via análises unicaracterísticas usuais.

6- Implicações

É importante deixar como consideração final, que este trabalho serviu apenas para ilustrar a metodologia do algoritmo SAEM aqui descrita como sendo um método alternativo de avaliação genética de curvas de crescimento. Outros trabalhos com um número maior de indivíduos e de comparação do algoritmo SAEM com outras metodologias de análise, como a análise conjunta Bayesiana, poderão ser avaliados para confirmar a eficiência definitiva desta metodologia.

7-Referências Bibliográficas

- ALBUQUERQUE; L.G. e FRIES, L.A. Modelos de regressão aleatória: possível aplicação na seleção de animais precoces. In: JOSAHKIAN, L. A. (ed.) Anais do 3o Congresso Brasileiro das Raças Zebuínas, ABCZ e SEBRAE. 29.11 - 02.12.1998. Uberaba. p 129-144, 1998.
- BLASCO A., PILES M., VARONA L., A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits, **Genetic Selection Evolution** v.35, p. 21–41, 2003.
- BRODY, S. **Bioenergetics and Growth**. Rheinhold Publishing, New York. 1945.
- CELEUX G., DIEBOLT J., The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, **Comp. Statist. Quat.** v.2 ,p. 73–82,1985.
- DELYON B., LAVIELLE M.,MOULINES E., Convergence of a stochastic approximation version of the EM algorithm, **Ann. Statist.**v 27,p. 94–128,1999.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM Algorithm (with discussion). **Journal of the Royal Statistics Society.** v. 39, p. 1 - 38, 1977.
- FARIA, C.U., MAGNABOSCO, C. U.,REYES, A.L.,LÔBO, R.B,BEZERRA, L.A.F. Inferência Bayesiana e sua aplicação na avaliação genética de bovinos da raça Nelore: revisão bibliográfica **Ciência Animal Brasileira** , v. 8, n. 1, p. 75-86, jan./mar. 2007
- FITZHUGH, H.A.Analysis of growth curves and strategies for altering their shapes. **Journal of Animal Science**, USA, v. 42, n.4, p.1036-1051, jul. 1976
- FORNI, S. **Análise da curva de crescimento de bovinos da raça nelore utilizando funções não lineares em análises bayesianas**. 2007. 75p. Tese (Doutorado em Zootecnia) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, SP.

- FREITAS, A.R.de. Curvas de crescimento na produção animal. **Revista Brasileira de Zootecnia**, v.34, n.3, p.786-795, 2005.
- FRIES, L.A. Precocidade, Precocidade e Precocidade. In: ENCONTRO DE PRODUTIVIDADE EM GADO NELORE, 1º, São Paulo, 7p., 1995.
- GARNERO, A. V.; MARCONDES, C. R.; BEZERRA, L. A. F.; OLIVEIRA, H. N.; LÔBO, H. N. Parâmetros genéticos da taxa de maturação e do peso assintótico de fêmeas da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v.57,n.5, p.652- 662,out. 2005.
- HENDERSON C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph - Canada, 1984. 462p
- JAFFRÉZIC,F., MEZA,C.,LAVIELLE, M.,FOULLEY,J.L.,Genetic analysis of growth curve using the SAEM algorithm. **Genetic Selection Evolution**, v.38, p. 583-600, aug/2006
- JUNGER, W.L., Estimação de parâmetros em relacionamento probabilístico de bancos de dados: uma aplicação do algoritmo EM para o reblink **Cadernos saúde coletiva**, Rio de Janeiro, v.14 , n.2 , p. 225 - 232, abr/jun 2006
- KUHN E., LAVIELLE M., Coupling a stochastic approximation version of EM with a MCMC procedure, **ESAIM Probability Statistic** v. 8,p. 115–131, 2004.
- KUHN E., LAVIELLE M.,Maximum likelihood estimation in nonlinear mixed effects models, **Computacion and Stastic**. Data Anal.v. 49, p. 1020–1038, 2005.
- MAZZINI, A.R. DE, MUNIZ, J. A., SILVA, F.F., AQUINO, L.H. Curva de crescimento de novilhos Hereford: heterocedasticidade e resíduos autorregressivos. **Ciência Rural**, v.35, n.2, p.422-427, mar-abr. 2005.
- MCCULLOCH C.E., Maximum likelihood algorithms for generalized linear mixed models, **Journal. Am. Statist. Assoc.** v.92, p. 162–170, 199
- MENDES, P.N.; MUNIZ, J.A.; SILVA, F.F.; MAZZINI, A.R.de.A.; SILVA, N.A.M Análise da curva de crescimento difásica de fêmeas Hereford por meio da função não linear de Gompertz. **Ciência Animal Brasileira**, v. 10, n. 2, p. 454-461, abr./jun. 2009
- MEYER, K. WOMBAT - **Digging deep for quantitative genetic analyses by restricted maximum likelihood** .In: World Congress on Genetics Appliedto Livestock Production, 8, 2006, Belo Horizonte. Proceedings...Belo Horizonte, 2006. I CD-ROM.
- MEZA,C. JAFFRÉZIC,F., FOULLEY,J.L., REML Estimation of Variance Parameters in Nonlinear Mixed Effects Models Using the SAEM Algorithm. **Biometrical Journal** , v. 49 ,n.6, p. 876–888, 2007.
- NOGUEIRA D. A; SÁFADI, T; BEARZOTI, E.; FILHO, J.S.S.B. Análise clássica e Bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e Agrotecnologia**, Lavras, Edição especial, p.1614-1624, dez. 2003.

OLIVEIRA, H.N. Análise genético-quantitativa da curva de crescimento de fêmeas da raça Guzerá. Universidade de São Paulo, Ribeirão Preto, SP, 73p., 1995 (Tese de Doutorado)

PANHARD, X.; SAMSON, A. Extension of the SAEM algorithm for nonlinear mixed models with 2 levels of random effects. **Biostatistics**, v.10, n.1, p. 121–135, 2009.

PONG-WONG, R.; HADJIPAVLOU, G. **A two-step approach combining the Gompertz growth with genomic selection for longitudinal data**. BMC proceedings, 2010, 4(suppl 1)S4

SAS Institute Inc. **Statistical Analysis System user's guide**. Version 9.1 ed. Cary: SAS Institute, USA, 2003.

SILVA, F. L **Curvas de crescimento e produtividade de vacas nelore e cruzadas, de diferentes tipos biológicos, em sistema de produção intensiva**. 2009. 89p. Dissertação (Mestrado em Zootecnia) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, SP.

SANTORO, R.S.; BARBOSA, S.B.P.; BRASIL, L.H.A.; SANTOS, E.S. Estimativas de parâmetros de curvas de crescimento de bovinos zebu, criados no estado de Pernambuco. **Revista Brasileira de Zootecnia**, Viçosa, MG, v.34, n.6, p.2262-2279, nov./dez. 2005.

SILVA, N.A.M.; MUNIZ, J.A.; SILVA, F.F.; AQUINO, L.H.; GONCALVES, T. M. Aplicação do método Bayesiano na estimação de curvas de crescimento de animais da raça Nelore. **Revista Ceres**, Viçosa, v.54, p.192-199, abr./maio 2008.

SILVA, N. A. M., LIMA, R. R., SILVA, F. F. E., AQUINO, L. H., MUNIZ, J. A. Modelo Hierárquico Bayesiano aplicado na avaliação genética de curvas de crescimento de bovinos de corte. **Arquivo Brasileiro de Medicina Veterinária e zootecnia.**, v.62, n.2, p.647-654, 2010

SILVA, N.A.M. da; AQUINO, L.H. de; SILVA, F.F; OLIVEIRA, A.I.G. de. Curvas de crescimento e influência de fatores não-genéticos sobre as taxas de crescimento de bovinos da raça Nelore. **Ciência e Agrotecnologia**, Lavras, v.28, n.3, p.647-654, maio./jun. 2004.

SILVA, F.F.; AQUINO, L.H.; OLIVEIRA, A.I.G. Estimativas de parâmetros genéticos de curva de crescimento de gado Nelore (*Bos indicus*). **Ciência e Agrotecnologia**, Lavras, v.26, p.1562-1567, 2002. Edição Especial.

SILVA, F.F. E ; AQUINO, L.H DE; OLIVEIRA, A.I.G. de. Influência de fatores genéticos e ambientais sobre as estimativas dos parâmetros de funções de crescimento em gado Nelore. **Ciência e Agrotecnologia**, v.25, n.5, p.1195-1205, 2001.

TALHARI, F.M.; ALENCAR, M.M.; MASCIOLI, A.S.; SILVA, M.A.; BARBOSA, P.F. Correlações genéticas entre características produtivas de fêmeas em um rebanho da raça Canchim **Revista Brasileira de Zootecnia**, Viçosa, v.32, n.4, p.880-886, jul./ago. 2003.

TEDESCHI, L.O.; BOIN, C.; NARDON, R.F.; LEME, P.F. Estudo da curva de crescimento de animais da raça guzerá e seus cruzamentos alimentados a pasto, com e sem suplementação. 1. análise e seleção das funções não-lineares **Revista Brasileira de Zootecnia**, Viçosa, v.29, n.2, p.630-637, mar./abr. 2000.

VARONA, L.; MORENO, C.; GARCIA CORTÉS, L.A; YAGUE, G.; ALTARRIBA, J. Two-step versus joint analysis of Von Bertalanffy function. **Journal of Animal Breeding and Genetics**, Berlin, v.116, p.331-338, jun. 1999