

Metodologias Aplicadas na Estimativa de Parâmetros Genéticos e Avaliações Genéticas de Caprinos e Ovinos

Raimundo Nonato Braga Lôbo

lobo@cnpq.embrapa.br

Pesquisador – Embrapa Caprinos
Fazenda Três Lagoas – Estrada Sobral Groaíras km 4
Caixa Postal D10 – 62011-970 – Sobral – CE

1. Introdução

O melhoramento genético das espécies de animais de interesse produtivo é feito, principalmente, a partir dos dados de controle zootécnico realizado nas propriedades. Estes dados são tomados em vários rebanhos, durante vários anos e diversos locais. A partir destas informações são realizadas avaliações genéticas dos indivíduos, ou seja, reprodutores, matrizes e produtos. Para a estimativa dos valores genéticos dos animais, é necessário a estimativa precisa de parâmetros genéticos, tais como, herdabilidade, repetibilidade e correlações genéticas, para as características consideradas na seleção.

Como o fenótipo do indivíduo é constituído pela parte genética, pela parte ambiental e muitas vezes pela interação entre ambas, o maior desafio destas estimativas está na separação destes componentes. Assim, a precisão das avaliações do valor genético depende da quantidade e da qualidade do banco de dados, da redução ou remoção das influências não genéticas nas informações analisadas, da incorporação de dados de parentes, do uso de modelos mistos apropriados e do método numérico adequado para a estimativa dos componentes de covariância e previsão do valor genético do animal.

No Brasil, estimativas de parâmetros genéticos e avaliações genéticas são escassas para caprinos e ovinos (Lôbo, 2002), ao contrário do que é observado para bovinos de corte e leite. Este quadro necessita ser modificado para que a exploração das espécies caprina e ovina seja inserida no competitivo mercado de alimentos e produtos de origem animal. Aqui serão discutidos os principais desafios, aspectos e métodos relacionados às estimativas de parâmetros genéticos e avaliações genéticas para caprinos e ovinos.

2. Modelos de Análises

Para remoção das influências não genéticas sobre as características utilizadas na seleção dos animais, as análises estatísticas devem ser feitas utilizando modelos apropriados, que são, geralmente, lineares uni ou multivariados, com efeitos fixos, aleatórios ou fixos e aleatórios, o que caracteriza os modelos mistos. Nos dias atuais a utilização do modelo animal tem sido uma constante, sendo o mais indicado nas avaliações genéticas e estimativas de parâmetros genéticos. Análises com este tipo de modelo, em caprinos e ovinos, são extremamente reduzidas (Lôbo, 2002). O modelo animal, além de permitir ajustes para os efeitos fixos, considera os dados do próprio animal e de seus parentes, possibilitando em uma única análise realizar a avaliação do próprio indivíduo e de outros relacionados a ele.

A qualidade de qualquer análise estatística depende do modelo assumido para descrever os dados. Um modelo deve representar adequadamente a natureza dos dados e refletir a biologia do problema (Schaeffer, 1993a). Há três tipos de modelo:

1. O modelo verdadeiro que descreve os dados perfeitamente, não deixando nenhum resíduo ou variação sem explicação; este nunca é conhecido exatamente;
2. O modelo ideal que é formulado pelo pesquisador, sendo o mais próximo possível do modelo verdadeiro; este é o que deve ser utilizado na análise, mas geralmente não há informação suficiente aplicá-lo;
3. O modelo operacional que é uma versão simplificada do modelo ideal, sendo aquele que o pesquisador utiliza em suas análises.

Um questionamento comum na formulação dos modelos de análises tradicionais é a distinção entre efeitos fixos e aleatórios. Sob, o ponto de vista da análise Bayesiana, não há distinção entre estes fatores. Efeitos fixos são aqueles em que todas suas classes são possíveis de ser observadas, como por exemplo sexo, classes de idade, ordem de lactação, sistema de manejo, etc. Como o número de fatores é pequeno ele pode ser repetido infinitamente inúmeras vezes em outros experimentos. Os efeitos aleatórios são aqueles cujos níveis são tomados aleatoriamente de uma população de níveis infinitamente grande. Estes fatores geralmente não podem ser repetidos em outro experimento. O efeito de animal deve ser considerado aleatório (Schaeffer, 1993a).

Os principais efeitos fixos que devem ser utilizados em modelos para análises de características de crescimento de caprinos e ovinos são aqueles inerentes a rebanho; raça ou grupo genético; sexo; ano e mês ou estação de nascimento; tipo de nascimento; idade ao parto ou ordem de parto da mãe e peso ou condição corporal da mãe ao parto, dentre outros. Para características relacionadas à produção de leite, destacam-se os efeitos de rebanho, raça ou grupo genético, ano e mês de parto, ordem de lactação, sistema de manejo, sistema e quantidade de ordenhas. Os efeitos de rebanho devem ser considerados uma vez que existem diferenças na qualidade e disponibilidade de forragem, no manejo e na variação genética dos rebanhos e, ainda, no nível de administração implementado pelos produtores. O potencial produtivo geralmente difere entre as raças e grupos genéticos, de forma que isto não pode ser negligenciado em uma análise que inclua diferente material genético. Ressalta-se que geneticamente somente pode ser considerada população aquele grupamento de indivíduos que partilham e trocam genes. Desta forma, geralmente, avaliações genéticas somente devem ser conduzidas dentro de cada raça ou grupo genético. Entretanto, tratando-se de avaliações de mestiços, oriundos de diversos cruzamentos, as características de cada grupo participante devem ser consideradas. A inclusão dos efeitos de ano, mês ou estação de nascimento ou parto nas análises é importante para a remoção das diferenças climáticas que afetam a disponibilidade de alimentos e o manejo, além das diferenças estacionais de administração do rebanho.

De maneira geral, os machos apresentam maior taxa de crescimento do que as fêmeas, sendo esta diferença atribuída a maior capacidade genética dos indivíduos do sexo masculino, possivelmente, devido a fatores hormonais. Animais oriundos de nascimentos múltiplos geralmente apresentam menor taxa de crescimento do que aqueles de nascimentos simples, principalmente devido à competição pelo alimento e a habilidade materna. Os fatores relacionados à ordem de parto ou idade da matriz e seu peso ou condição corporal ao parto são importantes pois refletem o estágio fisiológico da matriz. Fêmeas mais jovens, ainda em crescimento, produzem crias mais leves, devido ao menor desenvolvimento dos órgãos reprodutivos e menor irrigação do útero, com possível competição entre feto e mãe quanto aos nutrientes. Ainda, fêmeas de

primeira ordem de parto produzem menos leite do que aquelas de segunda ou maior ordem. Da mesma forma, sabe-se que, devido às deficiências na irrigação placentária que impedem maior passagem dos nutrientes, matrizes mais velhas também produzem crias mais leves, além de produzirem menor quantidade de leite. Obviamente que o sistema de manejo interfere nas respostas dos animais e deve ser considerado.

É presumível que quanto maior for a remoção da influência não genética, mais apurada será a comparação entre os animais por meio das avaliações genéticas. Nos dias atuais é comum a utilização de grupo contemporâneos nas análises estatísticas, que consiste em agrupar diversos efeitos em um único efeito, por exemplo, rebanho-ano-estação.

Para a determinação do melhor modelo de análise alguns critérios devem ser considerados. O coeficiente de determinação (R^2), o quadrado médio do resíduo e o erro de predição médio proposto por Goonewardene et al. (1981) são alguns destes critérios. Modelos com diferentes ordens de ajuste e efeitos aleatórios são comparados pelo teste da razão do logaritmo de verossimilhança (*log-likelihood ratio test* - LRT). Este teste somente permite comparações entre modelos aninhados e tende a favorecer àqueles com maior número de parâmetros (Olori et al., 1999; Meyer, 2000). Critérios de informação para formas de máxima verossimilhança restrita, tais como o Critério de informação de Akaike (AIC) e o Critério de Informação Bayesiano de Schwarz (BIC; Wolfinger, 1993), que impõem restrições de acordo com o número de parâmetros a ser estimado, também são utilizados.

3. Métodos de Estimação dos Componentes de Covariância

Segundo Henderson (1986), os componentes de covariância são importantes na obtenção da estatística F, na construção de índices de seleção, na análise de modelos mistos com vistas à predição linear do tipo Melhor Preditor Linear não Viciado (*Best Linear Unbiased Predictor* – BLUP), na estimativa dos parâmetros genéticos, fenotípicos e de meio ambiente, no planejamento de programas de melhoramento genético e na interpretação do mecanismo genético de características quantitativas.

São vários os métodos de estimação dos componentes de covariância para características múltiplas e dados não balanceados: Métodos I, II, III (Henderson, 1953) e IV de Henderson (Henderson, 1984); Método dos estimadores não viciados de norma mínima (*Minimum Norm Quadratic Unbiased Estimators* – MINQUE; Rao, 1971a); Método da estimação não viesada de mínima variância quadrática (*Minimum Variance Quadratic Unbiased Estimation* – MIVQUE; Rao, 1971b); Método da Máxima Verossimilhança (*Maximum Likelihood* – ML; Hartley & Rao, 1967); Método da Máxima Verossimilhança Restrita (*Restricted Maximum Likelihood* – REML; Patterson & Thompson, 1971); Método da falsa esperança (Schaeffer, 1986) e Método bayesiano da verossimilhança integrada (*Variance Estimation from Integrated Likelihoods*; Gianola & Foulley, 1990).

Verneque (1994) apresentou uma extensa revisão sobre estes métodos, que serão resumidos nos próximos itens.

3.1. Método I de Henderson

Um dos métodos mais simples de estimar os componentes de variância, aplicado aos procedimentos de análise de variância para dados balanceados, isto é, igual número de observações em todas as subclasses e sem covariáveis, ou desbalanceados em modelos completamente aleatórios. Produz estimadores não viciados e invariantes à

translação, ou seja não são afetados por mudanças nos efeitos fixos. Para estimar os componentes de variância, iguala-se às esperanças das formas quadráticas aos respectivos quadrados médios obtidos na análise de variância (ANOVA).

3.2. Método II de Henderson

Uma vez que o Método I somente poderia ser utilizado para modelos completamente aleatórios, foram desenvolvidos os Métodos II e III de Henderson, que podem ser utilizados para modelos mistos. O Método II é um procedimento imparcial e invariante à translação, fornece estimadores únicos para os componentes de covariância, mas não pode ser utilizado em modelos com interações entre efeito fixos e aleatórios ou aninhamentos do fator aleatório dentro de efeitos fixos. Para a obtenção do vetor solução, utiliza-se a inversa generalizada da matriz de coeficientes das equações dos quadrados mínimos para todos os fatores do modelo, de forma que diferentes vetores solução fornecem resposta única para os componentes de variância.

3.3. Método III de Henderson

Pode ser utilizado para qualquer modelo misto, entretanto requer mais tempo e maior capacidade de memória que os métodos anteriores. Utiliza as reduções nas somas dos quadrados, devido a submodelos do modelo completo, de forma que é chamado método de ajuste de constantes Yates. Fornece estimadores não viciados e invariantes à translação, entretanto estes estimadores não são únicos, uma vez que é possível definir ou obter maior número de reduções do que as necessárias. Pode fornecer estimativas mais precisas do que as obtidas pelos Métodos I e II.

3.4. Método IV de Henderson

No caso de grande volume de dados, é necessária a inversão da matriz de coeficientes de grande dimensão, o que torna o processo de estimação impraticável para os modelos anteriores. O Método IV de Henderson propõe o uso de uma inversa aproximada desta matriz de coeficientes. Sua base é a absorção dos efeitos fixos nas equações dos quadrados mínimos para os efeitos aleatórios. Esta absorção ajusta o lado direito das equações, que corresponde aos efeitos aleatórios, pelos efeitos fixos, o que permite que qualquer forma quadrática envolvendo um efeito absorvido seja invariante à translação.

3.5. Métodos MINQUE e MIVQUE

Até a década de 60 os métodos apresentavam apenas a imparcialidade e a invariância à translação como propriedades. Igualando-se os quadrados médios da ANOVA aos respectivos valores esperados obtém-se estimadores com estas propriedades e com mínima variância amostral. Entretanto, isto não é possível para dados desbalanceados, a não ser que se conheça as matrizes de variância e covariância dos fatores aleatórios e do erro, pelo menos de forma proporcional. Para contornar isto, no final da década de 60, novos métodos foram desenvolvidos. O método MIVQUE propunha derivar estimadores quadráticos, localmente melhores, invariantes à translação e imparciais, sob pressuposição de multinormalidade. O método baseia-se na estimação de funções quadráticas das observações e usa a restrição de que a matriz núcleo seja determinada, de forma que os estimadores tenham variância mínima. O procedimento MIVQUE requer os elementos inversos das equações do modelo linear misto para obtenção dos valores esperados das formas quadráticas, o que muitas vezes impossibilita o método para o caso de elevada ordem das equações.

Para quando as observações não seguem qualquer tipo de distribuição foi desenvolvido o método MINQUE ou método da estimação não viesada de norma mínima, que se baseia na estimação de funções quadráticas dos componentes de covariância, utilizando-se formas quadráticas das observações, com a restrição que a norma Euclidiana da matriz núcleo seja mínima.

Os métodos MIVQUE e MINQUE assumem que a matriz de covariância das observações é conhecida. Assim, a variância da forma quadrática para o MIVQUE ou a norma Euclidiana da matriz núcleo para o MINQUE só serão minimizadas se o valor preliminar da matriz de covariâncias das observações for o verdadeiro valor da população, o que produz estimadores imparciais, invariantes à translação e de variância ou norma mínima. Entretanto, na prática, a matriz de covariância não é conhecida, já que na verdade é o que se pretende estimar. Desta forma, estes métodos não são de variância ou norma mínima, mas são tão bons estimadores quanto mais próximo o valor inicial da matriz de covariâncias for do valor verdadeiro (Schaeffer, 1993b).

Sob normalidade os estimadores são equivalentes para ambos os métodos. Como na área biológica geralmente a multinormalidade é assumida, o método MIVQUE é mais utilizado e conhecido.

3.6. Método da Máxima Verossimilhança (ML)

Este método consiste na obtenção na função de verossimilhança que é representada por Λ ou função de densidade de probabilidade conjunta das observações, para um dado modelo de análise, em que se conhece a distribuição dos dados e existem parâmetros a serem estimados. As estimativas de ML para um conjunto de dados são valores numéricos dos parâmetros para o quais Λ é máximo. Por facilidades operacionais ou computacionais geralmente na prática se maximiza o valor do logaritmo da função de verossimilhança.

As principais vantagens deste método:

1. Produz estimadores de funções de estatísticas suficientes, consistentes, assintoticamente normais e eficientes;
2. Podem ser adotados em dados amostrais não aleatórios (Meyer, 1993);
3. Restrições para não negatividade nos componentes de variância ou dos autovalores das matrizes de covariância, ou ainda outras restrições no espaço paramétrico não causam dificuldades conceituais na aplicação de método.

Entretanto, o método apresenta as seguintes desvantagens:

1. Assume que a distribuição dos dados é conhecida. Geralmente, nas áreas biológicas, pressupõe-se que há distribuição multinormal na estimação dos componentes de covariância;
2. Produz estimadores viciados, uma vez que os efeitos fixos são tratados como se fossem conhecidos. A perda dos graus de liberdade resultante das estimativas dos efeitos fixos faz com que os estimadores obtidos pelo método não coincidam com os estimados pelo método ANOVA ordinário, mesmo para dados balanceados.

3.7. Método da Máxima Verossimilhança Restrita (REML)

Considerando-se apenas a parte da função de verossimilhança que independe dos efeitos fixos é possível remover o viés decorrente da perda dos graus de liberdade no ajuste dos efeitos fixos para estimativa dos componentes de covariância. Este procedimento foi chamado Método da Máxima Verossimilhança Restrita (REML), no qual cada observação é dividida em partes independentes, uma que se refere aos efeitos

fixos e outra que se refere aos efeitos aleatórios. O método REML produz componentes de covariância idênticos aos estimados nos métodos ANOVA, no caso de dados balanceados.

O método REML também requer que as observações tenham distribuição multinormal. Seus estimadores são invariantes à translação, mas são viciados, uma vez que se impõem restrições para não negatividade dos componentes de variância ou dos autovalores das matrizes de covariância. Entretanto, o vício é menor do que o observado para o método ML. No caso multivariado, se os parâmetros a serem estimados estiverem muito próximos ao limite do espaço de parâmetros o viés pode ser maior. Isto é comum quando as matrizes dos componentes de covariância apresentam determinante tendendo a zero. Como a aplicação das estimativas é de cunho prático, é preferível o pequeno viés do que a utilização de estimativas fora do espaço dos parâmetros (Henderson, 1984).

Os Métodos ML e REML diferem, principalmente, por que o primeiro utiliza a função de verossimilhança do vetor de observações ou o logaritmo desta função, enquanto o segundo usa esta função para um conjunto de contrastes de erros, com esperança nula, que representa as observações ajustadas para os efeitos fixos.

Tanto o Método ML como o Método REML utilizam procedimentos iterativos, ou seja, utilizam um processo de resolução, de uma equação, de um problema, mediante uma seqüência finita de operações em que o objeto de cada uma é o resultado da que a precede. Este último apresenta quatro formas de derivar as equações:

1. Utilizando as formas quadráticas utilizadas no Método MIVQUE e derivando novos valores esperados, assumindo que os valores preliminares são iguais aos verdadeiros valores dos parâmetros;
2. Partindo-se das equações derivadas pelo método ML;
3. Partindo-se da função de verossimilhança restrita e diferenciando-se para obtenção do máximo com respeito aos parâmetros desconhecidos;
4. Adotando-se o método livre de derivadas, que consiste em encontrar o máximo da função de verossimilhança restrita por meio de procedimentos de procura seqüencial ou linear.

3.8. Método da Falsa Esperança

Neste método, assume-se que os valores preliminares das razões de variância são iguais aos verdadeiros valores e falsas esperanças são tomadas. Esta aproximação é a mesma utilizada para derivar fórmulas pelo método REML com o algoritmo da maximização das esperanças (EM), a partir das equações derivadas do método MIVQUE. Schaeffer (1986) derivou um grupo de formas quadráticas cujas falsas esperanças não envolvem os elementos inversos das equações do modelo misto, com estimativas sempre positivas. Estes estimadores são invariantes à translação e viciados, mas com grande vantagem de serem obtidos com pouca dificuldade computacional. Estas formas quadráticas foram expandidas por VanRaden (1986), para o caso de parentesco entre animais. Este método não é capaz de considerar a seleção tipo refugio, como o REML, o que leva a produzir estimativas viciadas para os componentes de covariância (Schaeffer, 1993b).

3.9. Método Bayesiano da Verossimilhança Integrada (VEIL)

Os componentes de covariância são obtidos a partir de verossimilhanças integradas utilizando procedimentos bayesianos (Gianola & Foulley, 1990).

Máxima verossimilhança (ML), na terminologia bayesiana, é derivada da maximização da densidade conjunta “a posteriori” dos efeitos fixos (β) e dos componentes de covariância (V) dado o vetor de observações (y):

$$p(\beta, V | y) \propto p(y | \beta, V)$$

Se esta densidade conjunta for integrada em relação a β e em seguida a função de verossimilhança marginal for maximizada em relação aos componentes de covariância, têm-se os estimadores REML.

O método VEIL utiliza o mesmo conceito, integrando a verossimilhança marginal em relação a cada componente da matriz V, assumindo que outros valores destes componentes de variância tenham sido estimados (Schaeffer, 1993). As fórmulas resultantes são similares às dos métodos REML e ML, com a diferença de que são considerados os graus de liberdade utilizados para estimar β e cada um dos componentes de covariância.

O método VEIL também é iterativo.

Dentre os métodos apresentados, o método REML tem sido considerado o preferido para análise de dados desbalanceados na área de melhoramento animal. A preferência decorre principalmente das seguintes propriedades estatísticas que são desejáveis (Harville, 1977):

1. Os estimadores são funções de estatística suficientes, consistentes e assintoticamente normais e eficientes;
2. As aproximações são sempre bem definidas;
3. As restrições para não negatividade dos componentes de variância ou dos autovalores das matrizes de variância e covariância, ou outras restrições no espaço de parâmetros, não causam dificuldades conceituais na aplicação do método;
4. As soluções são quase sempre no espaço dos parâmetros;
5. As variâncias amostrais e o erro quadrático médio são menores do que aquelas obtidas por outros métodos que produzem estimadores não viciados;
6. Há redução ou eliminação do viés de pequenas amostras e ou decorrentes da seleção de animais (Meyer, 1983 e 1989).

4. Análises para Casos Particulares

Características que são consideradas função de alguma variável independente e contínua, como idade, necessitam de procedimentos eficientes e diferenciados para a realização de análises genéticas. Na área animal as principais características que consideram estas análises estão relacionadas à lactação de matrizes leiteiras e as taxas de crescimento. Os principais métodos para estas análises são a regressão aleatória, os polinômios ortogonais e os modelos de processos de caráter (*character process model*). Estes métodos foram avaliados por Jaffrézic & Pletcher (2000). Todos os três métodos são baseados na estimativa de verossimilhança, entretanto os polinômios ortogonais foram originalmente publicados como estimativas dos quadrados mínimos (Kirkpatrick et al, 1990).

Para esta discussão, será assumido, para um modelo aditivo, que o fenótipo observado para cada idade t pode ser decomposto como:

$$X(t) = \mu(t) + g(t) + e(t) + \varepsilon,$$

em que $\mu(t)$ é uma função não aleatória, a função média genotípica de $X(t)$, e $g(t)$ e $e(t)$ são funções aleatórias Gaussianas, que são independentes uma da outra e tem valor esperado igual a zero em cada idade. Elas representam os desvios genéticos e ambientais, respectivamente, dependentes da idade. Neste contexto, $e(t)$ é sempre referido com efeito ambiental permanente e ε é a variação residual, assumida normalmente distribuída com variância constante e desconhecida para toda idade.

O objetivo da análise é decompor a variação observada em $X(t)$ em seus componentes genéticos e ambientais pela estimativa de funções de covariância para $g(t)$ e $e(t)$. Uma função de covariância, $r(s,t)$, é uma função bivariada contínua que descreve a covariância entre duas idades, $r(s,t) = \text{Cov} [X(s), X(t)]$. Pela independência entre $g(t)$ e $e(t)$, a função de covariância fenotípica de $X(t)$ é dada por $P(s,t)$ como:

$$P(s,t) = G(s,t) + E(s,t),$$

em que $G(s,t)$ é a função de covariância genética e $E(s,t)$ é a função de covariância ambiental, que inclui também a variância residual. Estas funções são estimáveis pelos métodos ML e REML quando há dados de indivíduos aparentados.

Os **modelos de regressão aleatória** empregam formas paramétricas para funções não observadas. Embora tradicionalmente uma curva paramétrica média seja geralmente usada para estimar $\mu(t)$, isto não é essencial. Entretanto, os desvios individuais desta curva, ou seja, $g(t)$ e $e(t)$, são assumidos funções paramétricas do tempo, e polinômios geralmente são utilizados. Por exemplo, os desvios idade-dependente da média de uma população devido ao genótipo do indivíduo devem ser lineares no tempo, tal como:

$$G(t) = a_1 + a_2t,$$

em que a_i são coeficientes de regressão genética aleatória. Os coeficientes de regressão são efeitos aleatórios não observáveis, com um valor específico para cada indivíduo e assumidos com distribuição normal multivariada. Os desvios ambientais $e(t)$ são assumidos independentes dos efeitos genéticos e modelados similarmente.

Covariâncias genéticas e ambientais em função da idade são determinadas pelas variâncias e covariâncias entre os coeficientes de regressão.

O principal objetivo destes modelos é selecionar as funções paramétricas mais apropriadas para os desvios genéticos e de ambiente permanente. Em muitos casos as funções paramétricas estão aninhadas e a razão de verossimilhança pode ser utilizada. Desde que isto envolve o teste de significância dos parâmetros nos limites de seu espaço paramétrico, os testes estatísticos são misturas de distribuições de qui-quadrado (Stram & Lee, 1994).

Os **modelos de processo de caráter**, em contraste aos modelos de regressão aleatória, não tentam modelar as formas das funções $g(t)$ e $e(t)$. Em vez disto, o objetivo da análise é estimar os parâmetros para as próprias funções de covariância, ou seja, $G(s,t)$ e $E(s,t)$.

A função de covariância pode ser decomposta como:

$$G(s,t) = v_G(s)v_G(t)\rho_G(|s-t|),$$

em que $v_G(t)^2$ descreve o quanto da variância genética muda com a idade e $\rho_G(|s-t|)$ descreve a correlação genética entre duas idades. Não há restrições na forma de $v_G(\cdot)$, e é sempre modelado usando polinômios simples, ou seja, linear, quadrático, etc. É assumida correlação estacionária, ou seja, a correlação entre duas idades é assumida ser

função somente da distância temporal entre elas ($|s-t|$). Esta pressuposição geralmente é equivocada, entretanto prover razoável aproximação (Pletcher & Geyer, 1999). O benefício da correlação estacionária é que permite numerosas escolhas para $\rho(\cdot)$, todas que satisfazem diversos requerimentos teóricos (Pletcher & Geyer, 1999). Nunez-Anton (1998) e Nunez-Anton & Zimmerman (2000) propuseram métodos para considerar correlações não estacionárias.

Kirkpatrick & Heckman (1989) originalmente apresentaram o uso dos **polinômios ortogonais** como ferramenta não paramétrica para regularização prévia das estimativas das matrizes de covariâncias. Isto foi a primeira tentativa de formalizar a estimativa das funções de covariância no contexto genético. Como os modelos de processos de caráter, as formas dos desvios individuais idade-dependente não são consideradas, e os modelos para a estrutura da matriz de variância-covariância são o foco da atenção. Kirkpatrick & Heckman (1989) sugeriram que a função de covariância genética deveria ser representada como:

$$G(s,t) = \sum_{i=0}^m \sum_{j=0}^m \Phi_i(s) \Phi_j(t) k_{ij},$$

em que m determina o número de termos polinomiais usados no modelo, k_{ij} são os $m(m+1)/2$ parâmetros desconhecidos a serem estimados, ou seja, os coeficientes lineares de combinação) e Φ_i é o i -ésimo polinômio de Legendre. A função de covariância ambiental foi modelada similarmente. Meyer & Hill (1997) apresentaram um método para estimar as funções de covariância como descritas diretamente dos dados utilizando o método REML.

Como proposto originalmente, a aproximação por polinômios ortogonais é similar em intenção aos modelos de processos de caráter, e ambos diferem em princípio dos modelos de regressão aleatória. Nestes métodos de regressão aleatória, o desenvolvimento do modelo primário ocorre em nível de desvios individuais. O analista começa considerando o comportamento dos desvios individuais idade-específica. A estrutura da matriz de covariância resultante é consequência destes desvios. Para polinômios ortogonais e modelos de processo de caráter, a situação é reversa. O analista inicia considerando a estrutura da matriz de covariância e então a forma dos desvios individuais são consequências desta estrutura.

Características reprodutivas, como fertilidade ao parto, número de crias nascidas e desmamadas, e aquelas avaliadas visualmente, como tipo, conformação e desenvolvimento, apesar de serem importantes para a eficiência produtiva dos animais, ainda não têm sido utilizadas em programas de melhoramento animal devido suas baixas herdabilidades, geralmente obtidas por metodologias aplicadas a modelos lineares que assumem distribuição normal e que não são apropriadas para dados categóricos. Para análises destes dados, o mais indicado seria a utilização de modelos de limiar (*threshold models*) que assumem, subjacente ao fenótipo expresso de forma categórica, uma base genética e de ambiente com distribuição normal (Sousa et al., 2000).

Segundo Falconer & Mackay (1996), características que exibem fenótipo de distribuição discreta são normalmente chamadas quase-contínuas ou de limiar. Embora sejam classificadas dentro de uma ou várias categorias mutuamente exclusivas e exaustivas apresenta herança poligênica (Gianola, 1982).

Conforme Gianola (1982), o modelo de limiar assume que, conjuntamente ao fenótipo expresso de forma categórica, o qual frequentemente não apresenta distribuição normal, repousa uma postulada base genética e ambiental normalmente distribuída.

Valores na escala observável são ligados por uma escala subjacente, não observável, por intermédio de pontos de limiar entre categorias consecutivas.

Gianola & Foulley (1983) formalizaram um método para avaliação de características de limiar. A probabilidade de resposta em uma dada categoria segue uma integral normal com limites dependentes das variáveis fixas e aleatórias amostradas de uma distribuição conceitual com primeiro e segundo momentos conhecidos, a priori. A distribuição a priori e a função de verossimilhança são combinadas para fornecer uma densidade a posteriori, a partir da qual as inferências são feitas. Os parâmetros são estimados pela moda da distribuição a posteriori, cuja solução é tomada pela resolução de um sistema de equações não lineares.

5. Desafios para Avaliações em Caprinos e Ovinos

O principal desafio para a realização das estimativas de parâmetros genéticos e avaliações genéticas em caprinos e ovinos é a ausência de banco de dados em quantidade e em qualidade. A escrituração zootécnica ainda é deficiente mesmo nas instituições públicas. Por outro lado, essas anotações são, na maioria das vezes, inexistentes em nível de produtor. Aqueles que se dizem selecionadores, que criam animais registrados, limitam-se a fazer o registro genealógico dos animais junto as associações de criadores. Estas também não realizam controle produtivo, diferente do que é feito por associações de bovinos de corte e leite.

As poucas informações disponíveis, fruto do trabalho isolado de algumas instituições e produtores, não produzem estimativas robustas e confiáveis, face ao mecanismo de anotação e formação dos bancos de dados. Em geral, o maior entrave é a identificação dos animais, o que compromete a estruturação da matriz de parentesco, imprescindível para o uso nas metodologias utilizadas no processo de estimativas de parâmetros genéticos e dos valores genéticos dos animais.

Dentro de uma visão otimista, o estado atual de ausência das informações abre caminho para o estímulo e incentivo a uma escrituração zootécnica eficiente, abrangente e bem conduzida. É possível verificar as falhas observadas em outros setores produtivos e tentar não incorrer nos mesmos erros.

Outro ponto que merece destaque é a falta de conectabilidade entre os rebanhos. A teoria dos índices de seleção somente permite comparações entre animais que compartilhem o mesmo espaço e tempo, ou seja, mesmo grupo contemporâneo. O método dos quadrados mínimos e seus derivados permitem comparações livres das dimensões espaço-tempo, desde que os dados possuam uma boa estrutura e sejam conectados. O problema é que “conectabilidade” é definida, em estatística, como uma característica discreta, de qualidade com respeito a cada observação. Assim, não existe um procedimento padrão para avaliar se uma determinada estrutura de dados é capaz de bem explorar ou não o potencial de uma determinada metodologia de avaliação genética (Fries & Roso, 1997). A metodologia dos modelos mistos de Henderson fornece soluções sempre, para qualquer estrutura de dados. Não existe qualquer exigência mínima de conectabilidade, nem um padrão mínimo de referência e nem uma forma universal de medir a quantidade de conexões dentro de um conjunto de dados. Esta situação ainda é entendida como uma das vantagens da metodologia. Entretanto, Fries & Roso (1997) afirmaram que como o conhecimento de que as chamadas propriedades genéticas das soluções das equações dos modelos mistos na verdade são apenas as partes visíveis, reconhecida por todos e resultantes dos somatórios das restrições impostas, a realização de análise de conectabilidade antes das avaliações genéticas se

torna, a cada dia, mais e mais recomendáveis e logo deverão chegar ao nível de obrigatoriedade.

Com a reduzida utilização das biotécnicas de reprodução, principalmente a inseminação artificial, com o uso de reprodutores em monta natural exclusivamente dentro de uma mesma propriedade, a conectabilidade entre rebanhos caprinos e ovinos é reduzida, o que impede avaliações genéticas mais abrangentes, limitando-as para dentro de cada rebanho. Assim, a avaliação de reprodutores em diversos ambientes e sob diferentes condições de manejo, que é importante para o melhoramento genético dos caprinos e ovinos, em todo território brasileiro, fica comprometida.

6. Considerações Finais

- As metodologias de análises existentes são robustas e tem permitido avanços nos programas de melhoramento animal de diversas espécies em todo mundo. No Brasil, estas técnicas são utilizadas em larga escala, merecendo destaque a evolução do nosso gado de corte, devida principalmente a contribuição da genética quantitativa. Em caprinos e ovinos, a utilização destas metodologias é ainda muito tímida e discreta.
- O passo inicial para a mudança da situação atual é o estímulo à escrituração zootécnica junto aos produtores e sua integração com as instituições de pesquisa e ensino.
- O processo de identificação única e exclusiva de animais, através da utilização de processos em duplicata, como o uso de brincos ou medalha associado à tatuagem, ou até mesmo a utilização de chips eletrônicos, permitirá a formação de pedigrees com menores falhas, o que auxiliará na confiabilidade das estimativas.
- Um marco para esta mudança foi o lançamento, em 2003, pela Embrapa Caprinos, do Programa de Melhoramento Genético de Caprinos e Ovinos de Corte (GENECOC).
- Outras iniciativas, principalmente para o sistema de produção de leite, estarão sendo lançadas em breve, também por esta unidade.

7. Referências Bibliográficas

- FALCONER, D.S., MACKAY, T.F.C. Introduction to quantitative genetics. 4.ed. London: Longman, 1996.
- FRIES, L.A., ROSO, V.M. Conectabilidade em avaliações genéticas de gado de corte: uma proposta heurística. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 34., 1997, Juiz de Fora, MG. *Anais...*Juiz de Fora: Sociedade Brasileira de Zootecnia, 1997.v.3, p.159-161.
- GIANOLA, D. Theory and analysis of threshold characters. *J Anim. Sciences*, v. 54, p.1079-1080, 1982.
- GIANOLA, D., FOULLEY, J.L. Sire evaluation for ordered categorical data with a threshold model. *Génétique, Sélection, Évolution*, v.15, p.201-224, 1983.
- GIANOLA, D., FOULLEY, J.L. Variance estimation from integrated likelihood (VEIL). *Génétique, Sélection, Évolution*, v.22, p.403-417, 1990.

- HARTLEY, H.O., RAO, J.N.K. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, v.54, n.1., p.93-108, 1967.
- HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, Chicago, v.72, n.358, p.320-339, 1977.
- HENDERSON, C.R. Applications of linear models in animal breeding. Ontario, University of Gueph, 1984. 462p.
- HENDERSON, C.R. Estimation of variance and covariance components. *Biometrics*, v.9, p.226-252, 1953.
- HENDERSON, C.R. Recent developments in variance and covariance estimation. *Journal of Animal Sciences*, v.63, p.208-216, 1986.
- JAFFRÉZIC, F., PLETCHER, S.D. Statistical models for estimating the genetic the genetic basis of repeated measures and other function-valued traits. *Genetics*, v.156, p.913-922, 2000.
- KIRKPATRICK, M., HECKMAN, N. A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of Mathematical Biology*, v.27, p.429-450, 1989.
- KIRKPATRICK, M., LOFSVOLD, D., BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, v.124, p.979-993, 1990.
- LÔBO, R.N.B. Melhoramento genético de caprinos e ovinos: desafios para o mercado. Sobral, CE: Embrapa Caprinos, 2002. 36p. (Embrapa Caprinos. Documentos, 39).
- MEYER, K. DFREML – Version 2.1.09. User notes. Armidale, University of New England, 1993. 97p.
- MEYER, K. Maximum likelihood procedures for estimating genetic parameters for later lactations in dairy cattle. *Journal of Dairy Science*, Champaign, v.66, n.9, p.1988-1997, 1983.
- MEYER, K. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livestock Production Science*, v.65, p.19-38, 2000.
- MEYER, K. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genétique, Sélection, Évolution*, Paris, v.21, p.317-340, 1989.
- MEYER, K., HILL, W.G. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by Restricted Maximum Likelihood. *Livestock Production Science*, v.47, p.185-200, 1997.
- NUNEZ-ANTON, V. Longitudinal data analysis: non-stationary error structures and antedependent models. *Appl. Stochastic Models Data Anal*, v.13, p.279-287, 1998.
- NUNEZ-ANTON, V., ZIMMERMAN, D.L. Modeling non-stationary logitudinal data. *Biometrics*, v.56, n. 3, p.699-705, 2000.
- OLORI, V.E., HILL, W.G., McGUIRK, B.J., BROTHERSTONE, S. Estimating variance components for test day milk records by restricted maximum likelihood with random regression animal model. *Livestock Production Science*, v.61, p.53-63, 1999.
- PATTERSON, H.D., THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, v.58, p.545-554, 1971.

PLETCHER, S.D., GEYER, C.J. The genetic analysis of age- taken. y and y' are the means of the correlation values dependent traits: modeling a character process. *Genetics*, v.153, p.825–833, 1999.

RAO, C.R. Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, v.1, p. 257-275, 1971a.

RAO, C.R. Minimum variance quadratic unbiased Estimation of Variance Components. *Journal of Multivariate Analysis*, v.1, p.445-456, 1971b.

SCHAEFFER, L.R. Estimation of variance and covariances within the allowable parameter space. *Journal of Dairy Science*, v.69, n.1, p.187-194, 1986.

SCHAEFFER, L.R. Linear models and computing strategies in animal breeding. Ontario, University of Guelph. 1993a. 215p.

SCHAEFFER, L.R. Variance component estimation methods. Ontario, University of Guelph, 1993b. 113p.

SOUSA, W.H., PEREIRA, C.S., BERGMANN, J.A.G, SILVA, F.L.R. Estimativas de componentes de variância e de parâmetros genéticos para características de reprodução por intermédio de modelos lineares e de limiar. *Revista Brasileira de Zootecnia*, v.29, n.6, p.2237-2247, 2000.

STRAM, D.O., Lee, J.W. Variance components testing in the longitudinal and mixed effects model. *Biometrics*, v.50, p.1171–1177, 1994.

VanRADEN, P.M. Computational strategies for estimation of variance components. Ames, 1986. 112p. (PhD – Iowa Sate University).

VERNEQUE, R.S. Procedimentos numéricos e estimação de componentes de covariância em análise multivariada pelo método da máxima verossimilhança restrita – modelos mistos aplicados ao melhoramento animal. Piracicaba, SP: ESALQ – USP, 1994. 157p. Tese de Doutorado.

WOLFINGER, R.D. 1993. Covariance structure in general mixed models. *Commun. Statist.*, 22B, p.1079-1106, 1993.