

AVALIAÇÃO GENÉTICA DE ANIMAIS PARA CARACTERÍSTICAS CATEGÓRICAS – UMA REVISÃO

William José Ferreira¹

FERREIRA¹, W. J. Avaliação genética de animais para características categóricas – uma revisão. *Arq. Ciênc. Vet. Zool. Unipar*, Umuarama, v. 11, n. 2, p. 121-128, jul./dez. 2008.

RESUMO: Existem muitas características de interesse biológico ou importância econômica, cuja herança é multifatorial, mas cuja distribuição é descontínua. Características deste tipo parecem estar fora da conjuntura da genética quantitativa, porém, quando submetidas a análise genética, elas são encontradas como se estivessem sob a influência de muitos genes, à semelhança das características métricas. A maioria das avaliações genéticas de animais têm sido realizadas com base na pressuposição de distribuição contínua da característica estudada. No entanto, muitas das características de importância econômica têm uma distribuição descontínua do fenótipo e, neste caso, a metodologia aplicada às características que apresentam distribuição contínua não é a mais indicada. Por isso, objetivou-se, com este estudo, realizar uma breve revisão sobre metodologias de avaliação genética de animais, para características categóricas. Foi estudada a aplicação de métodos lineares, não-lineares e do modelo de limiar. Com base na revisão de literatura realizada concluiu-se que, para características categóricas, a utilização do modelo de limiar, que considera a natureza multinomial das características, é mais acurada para utilização em avaliações genéticas de animais.

PALAVRAS-CHAVE: Componentes de variância. Método REML. Modelo animal. Modelo de limiar. Parâmetros genéticos.

ANIMAL GENETIC EVALUATION FOR CATEGORICAL CHARACTERISTICS – A REVIEW

ABSTRACT: There are a number of characteristics implying in biological interesting or economic importance whose legacy is multi factorial, but whose distribution is discontinuous. Such characteristics seem to be apart from the quantitative genetics conjuncture; however, when submitted to genetics analysis, they are as if they were under the influence of many genes influence, as the metric characteristics. Most of animal genetics evaluations have been conducted based on the presupposition of ongoing distribution of the characteristic studied. Nevertheless, most of the economically important characteristics present discontinuous distribution of phenotype, and, in that case, the methodology applied to the characteristics presenting continuous distribution is no longer recommended. Therefore, the purpose of this study was to carry out a brief review on animal genetics assessment methodologies for categorical characteristics. The application of linear and non-linear methods and the threshold model were studied. Based in this literature review it was conclude that, for categorical characteristics, the use of the threshold model, which considers the multinomial nature of the characteristics is the most accurate to be used for animal genetics evaluation.

KEYWORDS: Variance components. REML method. Animal model. Threshold model. Genetics parameters.

EVALUACIÓN GENÉTICA DE ANIMALES PARA CARACTERÍSTICAS CATEGÓRICAS – UNA REVISIÓN

RESUMEN: Existen muchas características de interés biológico o importancia económica, cuya herencia es multifactorial, pero cuya distribución es discontinua. Características de este tipo parecen estar fuera de la coyuntura de la genética cuantitativa, sin embargo, cuando sometidas a análisis genética, ellas son encontradas como si estuviesen bajo la influencia de muchos genes, a semejanza de las características métricas. La mayoría de las evaluaciones genéticas de animales han sido realizadas con base en la presuposición de distribución continua de la característica estudiada. Sin embargo, muchas de las características de importancia económica tiene una distribución discontinua del fenotipo y, en este caso, la metodología aplicada a las características que presentan distribución continua no es la más indicada. Por eso, se buscó, con este estudio, realizar una breve revisión sobre metodologías de evaluación genética de animales, para características categóricas. Fue estudiada la aplicación de métodos lineares, no lineares y del modelo de limiar. Con base en la revisión de literatura realizada se concluyó que, para características categóricas, la utilización del modelo de limiar, que considera la naturaleza multinomial de las características, es más perfeccionada para utilización en evaluaciones genéticas de animales.

PALABRAS CLAVE: Componentes de variancia. Método REML. Modelo animal. Modelo de limiar. Parámetros genéticos.

¹ Tecnólogo em Processamento de Dados. Especialista em Matemática e Estatística. Mestre e Doutor em Zootecnia. Professor e Diretor Geral da Faculdade do Noroeste de Minas (FINOM). Rodovia MG 188 KM 167, s/nº - Bairro Fazendinha - Caixa Postal Nº 201 - Paracatu-MG - CEP: 38600-000. Telefax: (38) 3671-2454. E-mail: finom@finom.edu.br. Site: www.finom.edu.br

Introdução

Existem muitas características de interesse biológico ou importância econômica, cuja herança é multifatorial, mas cuja distribuição é descontínua. Por exemplo, resistência a doenças, um caráter expresso em termos de sobrevivência ou morte sem intermediários; tamanho de leitegada que, em mamíferos, produz usualmente 1 indivíduo de cada vez, mas em alguns casos dois ou três; a presença ou ausência de qualquer órgão ou estrutura.

Segundo Falconer (1987), características deste tipo parecem estar, à primeira vista, fora da conjuntura da genética quantitativa. Quando submetidas à análise genética, elas são encontradas como estando sob a influência de muitos genes, à semelhança das características métricas. Por esta razão, elas têm sido consideradas como aquelas que apresentam “variações quase contínuas”.

Um ponto importante para o entendimento da herança de tais caracteres, apoia-se na idéia de que a característica tem uma continuidade básica, com um limiar que impõe uma descontinuidade sobre a expressão visível da característica (Figura 1). A variação é de origens genética e ambiental e poderia ser entendida como a concentração de alguma substância ou como a velocidade de um processo de desenvolvimento, que poderia, em princípio, ser medida e estudada como um caráter métrico, no sentido ordinário. A medida hipotética desta variação é supostamente feita numa escala que torna sua distribuição normal, e a unidade de medida é o desvio-padrão da distribuição. Isso fornece o que pode ser chamado de “escala base”. Tem-se, agora, duas escalas para a descrição dos valores fenotípicos: a escala base, que é contínua e a escala visível, que é descontínua. As duas são conectadas pelo limiar ou ponto de descontinuidade. Este é um ponto na escala contínua que corresponde à descontinuidade da escala visível.

O conceito de limiar pode ser ilustrado por uma característica, cuja expressão visível pode tomar apenas duas formas, tais como vivo versus morto; presente versus ausente. Indivíduos, cujos valores fenotípicos, na escala base, excedem o limiar, aparecerão numa classe visível enquanto os indivíduos abaixo do limiar (*threshold*), aparecerão em outra.

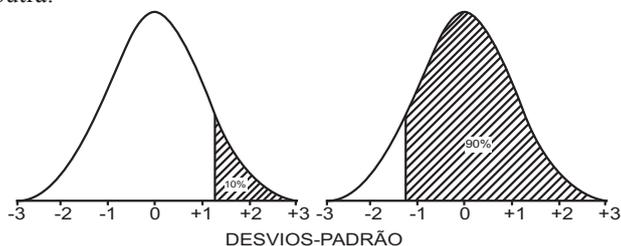


Figura 1: Desvio-padrão de uma característica de limiar com duas escalas visíveis.

Em cada uma delas, as linhas verticais marcam o limiar entre as duas classes fenotípicas, uma das quais é “hachuriada”. Na população representada do lado esquerdo a característica tem uma incidência de 10%; e na do lado direito, uma incidência de 90%

Numa escala visível, os indivíduos podem ter somente dois valores; zero ou um, codificados para presença ou ausência. Grupos de indivíduos, contudo, tais como famílias

ou a população como um todo, podem ter qualquer valor, na proporção ou porcentagem de indivíduos, em uma ou outra classe. Isso pode ser referido como incidência da característica. Susceptibilidade à doença, por exemplo, pode ser expressa como porcentagem de mortalidade, na população ou família. A incidência é bem adequada para uma descrição da população ou grupo, mas a escala de porcentagem, sobre a qual a incidência é expressa, é inapropriada para algumas finalidades, porque, na escala de porcentagem, as variâncias diferem de acordo com a média. A interpretação de análises genéticas de características de limiar é, portanto, facilitada pela transformação de incidência para valores numa escala base. A transformação é facilmente realizada por referência a uma tabela de probabilidade da curva normal.

Na área de melhoramento genético animal, os programas de seleção procuram metodologias que possam permitir a obtenção de avaliações genéticas cada vez mais precisas, o que resultará em maiores ganhos genéticos por geração. A maioria das avaliações de animais têm sido realizadas com base na pressuposição de distribuição contínua da característica estudada. No entanto, como muitas das características de importância econômica têm uma distribuição descontínua do fenótipo a metodologia aplicada às características que apresentam distribuição contínua não é a mais indicada.

Assim, objetivou-se, com este estudo, realizar uma breve revisão sobre metodologias de avaliação genética de animais, para características categóricas.

Revisão de Literatura

1. Análise de Características Categóricas em Melhoramento Animal

Em melhoramento genético aplicado à produção animal, a maioria das características de interesse econômico sobre as quais se pretende atuar têm uma distribuição fenotípica contínua. Entretanto, algumas outras de grande importância não podem ser associadas a uma distribuição contínua, tendo uma expressão fenotípica em poucas categorias, exclusivas mutuamente, como por exemplo: dificuldade de parto, viabilidade, número de animais nascidos, susceptibilidade a certas enfermidades, algumas características morfológicas ou de tipo, entre outras.

Freqüentemente, tem-se utilizado aproximações lineares (mínimos quadrados) para ajustar variáveis discretas (CAÑON; FUENTE, s.d.).

Um procedimento, considerado a base para os estudos e progressos obtidos no melhoramento genético, é de o Grizzle et al. (1969), que consiste em uma aproximação dos métodos de mínimos quadrados generalizados, freqüentemente denominado método GSK e que estes autores aplicaram a modelos fixos. Koch & Reinfurt (1971), expandiram o modelo de Grizzle et al. (1969) aos modelos mistos.

A maior vantagem que os métodos lineares têm apresentado é o fato de se poder utilizar os programas de análises já disponíveis e a grande flexibilidade que se tem quando são utilizados tais programas.

Thompson (1979) propõe um procedimento não-linear para classificar touros, no caso em que a variável estudada se ajusta a uma distribuição binomial, utilizando um modelo linear generalizado, cujas equações são similares às

expressas por Henderson (1973), para o modelo linear misto e uma variável contínua.

Gianola & Foulley (1983) desenvolveram um modelo de predição não-linear, supondo a existência de uma variável subjacente contínua e cujas soluções seriam semelhantes às obtidas pelo método proposto por Thompson (1979).

1.1. Delineamento Geral do Problema

Em melhoramento genético animal, ao se realizar a seleção para uma ou várias características, deve-se escolher aqueles indivíduos que serão os pais da próxima geração. A seleção dos reprodutores deve ser de acordo com o seu valor genético. O valor genético não pode ser observado ou medido diretamente, por isso a seleção deve basear-se em variáveis que passam ser medidas e estatisticamente relacionadas com o valor genético.

Observa-se nos trabalhos da área de melhoramento genético animal que, selecionando-se por truncamento, os candidatos com maiores valores de $E(u/y)$, sendo que u é o valor genético do indivíduo e y a variável observável, o valor esperado de u para os indivíduos selecionados maximiza y , havendo, portanto, ganho genético, supondo que os pares (u, y) são amostras independentemente e identicamente distribuídas, com uma distribuição normal multivariada. Henderson (1963) particularizou esta conclusão devido ao fato de que em melhoramento animal é difícil de se atender às restrições de independência e igualdade da distribuição das amostras, porque, se $\begin{bmatrix} u \\ y \end{bmatrix} \sim N(\mu, \Sigma)$, a probabilidade de ordenar corretamente um par u_i, u_j se maximiza ao fazê-lo de acordo com a razão $E(u/y)$ e $E(u/y)$.

A utilização do Melhor Preditor Linear Não-Viesado (BLUP) descrito por Henderson (1973) é plenamente justificada quando $E(u/y)$ é linear nas observações. Considerando-se que a probabilidade é discreta e não se ajusta a uma distribuição normal, a $E(u/y)$ não é linear em y e, portanto, o procedimento BLUP não tem condições de classificar os indivíduos corretamente. Como conseqüência, tem-se proposto alternativas ao modelo linear, como são os modelos não-lineares (THOMPSON, 1979; GIANOLA & FOULLEY, 1983; FOULLEY & GIANOLA, 1984) ou modelos lineares para as observações transformadas (GIANOLA, 1980a,b).

1.2. Aplicação de Métodos Lineares

Schaeffer & Wilton (1976) utilizaram, para prever o mérito genético de touros, para a característica facilidade de parto, o método proposto por Grizzle et al. (1969), de mínimos quadrados generalizados, aplicando um modelo misto às variáveis discretas, com as seguintes suposições: cada subpopulação (combinação de níveis das variáveis explicativas) é uma amostra aleatória de uma mesma população. Sob esta condição, o procedimento BLUP (HENDERSON, 1973) é uma modificação do método de mínimos quadrados ordinários para prever diferenças entre efeitos aleatórios. As soluções do método proposto por Grizzle et al. (1969) tem as propriedades BAN (Best Asintotic Normal).

Tong et al. (1976) utilizaram a metodologia proposta por Schaeffer & Wilton (1976), para avaliar touros da raça Charolês para característica facilidade de parto. Pollak & Freeman (1976), aplicaram esta mesma metodologia à ca-

racterística distocia. Um trabalho similar fizeram Berger & Freeman (1978), que analisaram três modelos mistos para encontrar o BLUP do mérito genético de touros para dificuldade de parto.

O modelo linear aditivo proposto por Schaeffer & Wilton (1976), considera que a variável estudada, ou seja, a variável que se expressa como função linear de alguns parâmetros é a pontuação que se associa a cada uma das categorias, em que se divide o caráter de interesse.

Quaas & Van Vleck (1980) consideraram cada uma das categorias, às quais uma população pode pertencer, como uma característica diferente e estimaram a probabilidade de que uma observação futura pertença a uma determinada categoria, utilizando BLUP. Um procedimento similar a este havia sido utilizado para prever a probabilidade de que uma descendência futura de um touro pertença a uma determinada categoria, para características de tipo (VAN VLECK & KARNER, 1980). A metodologia BLUP tem sido utilizada, freqüentemente, para classificar reprodutores para características de tipo e conformação (SCHAEFFER & BURNSIDE, 1974; SCHAEFFER et al., 1978).

Os métodos lineares apresentados, que são aplicados às características discretas, apresentam certos problemas metodológicos e conceituais, até mesmo colocando em dúvida a sua utilização (THOMPSON, 1979; GIANOLA, 1982), sendo os mais importantes citados em seguida:

1) A associação das pontuações para as diferentes respostas possíveis são arbitrárias (GRIZZLE et al., 1969; BERGER & FREEMAN, 1978) e, quando se supõe a existência de uma variável subjacente contínua, a herdabilidade (h^2) na escala real não é invariante em relação a estas pontuações (GIANOLA, 1982). Entretanto, é possível, mediante um delineamento apropriado, escolher um conjunto de pontuações que são mais herdáveis que outros. Snell (1964) e Gianola & Norton (1981) descreveram procedimentos com propriedades ótimas do ponto de vista estatístico e genético, respectivamente.

2) As soluções obtidas por meio do modelo linear misto não levam em consideração a restrição de que a soma das probabilidades de cada categoria dentro de uma população seja igual a 1, pois, ao se restringir o espaço de estimação, é possível obter um preditor com variância amostral menor.

3) A variância na escala real (supondo um modelo de limiar) não é constante de uma população a outra, uma vez que é função de $p_i/(1-p_i)$, sendo i uma população, e este produto pode variar muito entre populações.

1.3. Aplicação de Métodos Não-Lineares

Gianola (1980a,b) propõe um modelo em que as probabilidades se transformam mediante uma função logarítmica ($\log p_i/(1-p_i)$), e esta função se expressa como uma combinação linear das variáveis explicativas. O método de estimação que se utiliza é o de mínimo- X^2 , que fornece estimadores BAN; ou ainda, o método da máxima verossimilhança, maximizando a função de verossimilhança. Este segundo procedimento tem a vantagem teórica, pelo menos, de fornecer estimadores com variância mínima. O inconveniente é a necessidade de se utilizar um procedimento iterativo para obter as soluções, o que implica em um incremento das

necessidades computacionais, quando o número de classes e de efeitos aleatórios é grande, como ocorre em melhoramento animal.

Outro método que transforma as probabilidades foi proposto por Wright (1934), denominado posteriormente de "probit", definido como a transformação da inversa da probabilidade, definido como: $v_i = \Phi^{-1}(p_i)$, onde Φ representa a função de distribuição de uma normal.

Thompson (1979) coloca em dúvida a validade dos modelos lineares para se ajustar variáveis binomiais e, em relação aos problemas de avaliação de touros, propôs a substituição da variável estudada y , das equações normais, por $X\beta + Zu + [y - \Phi(X\beta + Zu)]W$, equações que não são lineares em y .

Posteriormente, Gianola & Foulley (1983) desenvolveram um método de avaliação de touros para características discretas ordenadas, com base na pressuposição de um modelo de limiar que supõe a existência de uma variável subjacente contínua e que a diferença entre limiares é fixa. O procedimento desenvolvido por Gianola & Foulley (1983) é uma aproximação Bayesiana que envolve o cálculo da moda do logaritmo de uma distribuição posterior (combinação da distribuição inicial e de verossimilhança).

Equações similares foram obtidas por Harville & Mee (1984), para um modelo misto, que prediz o valor de uma variável discreta ordenada. Este método também baseia-se em um modelo de limiar, que supõe que o valor observado é determinado pelo valor de uma variável subjacente não observável, contínua e que pode ajustar-se a um modelo linear.

É interessante destacar ainda, alguns trabalhos que descrevem procedimentos não-lineares, para diferentes aplicações.

Foulley et al. (1983) apresentaram um método de predição de valores genéticos de touros para duas características conjuntamente, uma das quais é discreta e a outra contínua, seguindo o procedimento de maximizar uma função "a posteriori".

Foulley e Gianola (1984) obtiveram as equações para avaliar touros para duas características discretas binomiais, calculando igualmente a moda da distribuição final conjunta.

Perez-Enciso et al. (1993) apresentaram uma metodologia que implementa um modelo animal, considerando-se um modelo de limiar e que permite realizar análises de duas características. Esta metodologia foi utilizada para prever valores genéticos para taxa de ovulação e sobrevivência embrionária.

Hoeschele et al. (1995) desenvolveram uma metodologia para predição de valores genéticos em análises de características múltiplas e com modelo animal, para uma característica categórica e número "ilimitado" de variáveis contínuas. A característica categórica pode ser expressa por qualquer número de categorias. Nesta metodologia, um modelo genérico, incluindo efeito materno, efeito permanente de ambiente ou outros efeitos aleatórios, pode ser usado, para cada característica, bem como modelos diferentes para cada característica e observações "perdidas" para algumas características.

1.3.1. Descrição do Modelo de Limiar (Threshold-Model)

O modelo de limiar foi proposto por Gianola & Foulley (1983) e utiliza o programa computacional desenvolvido por Misztal et al. (1989), com base no conceito de Dempster & Lerner (1950) e Falconer (1965). Este conceito assume que o processo de resposta está relacionado a uma variável contínua subjacente, normalmente distribuída, chamada Liability ("Predisposição") ℓ e a um conjunto de limiares fixos, $\delta' = [\delta'_1 < \delta'_2 \dots < \delta'_{m-1}]$, que divide a linha real dentro de m intervalos correspondendo às categorias de respostas, com $\delta'_0 = -8$ e $\delta'_m = +8$. A distribuição de ℓ no contexto do modelo multifatorial da genética quantitativa pode ser assumida como normal (DEMPSTER & LERNER, 1950; GIANOLA, 1982). Portanto, esta variável é considerada como o resultado da combinação linear de pequenos efeitos a partir de alelos em um grande número de loci, mais os componentes aleatórios e de ambiente.

De acordo com o proposto por Gianola & Foulley (1983), para análises de dados categóricos, os dados são organizados em uma tabela de contingência $s \times m$, como a esquematizada a seguir (quadro 1), onde as s linhas representam indivíduos ou combinações de níveis de variáveis explanatórias e as m colunas indicam categorias de respostas ordenadas mutuamente, exclusivas e exaustivas.

Nesse quadro, n_{jk} , são os números de unidades experimentais na k ésima categoria de resposta ($k=1, 2, \dots, m$) sob as condições da j ésima linha ($j=1, 2, \dots, s$). Totais de linhas, n_j ($j=1, \dots, s$) são considerados como fixos. As variáveis aleatórias de interesse são $n_{j1}, n_{j2}, \dots, n_{jk}, \dots, n_{jm}$, para $j=1, \dots, s$.

Quadro 1 - Dados categóricos ordenados, arranjados como uma tabela de contingência ($s \times m$), segundo Gianola e Foulley (1983)

Linha ¹	Categoria da resposta ²				Total fixo ³
	1	2 . . .	k	...m	
1	n_{11}	$n_{12} \dots$	n_{1k}	$\dots n_{1m}$	$n_{1.}$
2	n_{21}	$n_{22} \dots$	n_{2k}	$\dots n_{2m}$	$n_{2.}$
.
.
j	n_{j1}	$n_{j2} \dots$	n_{jk}	$\dots n_{jm}$	$n_{j.}$
.
.
S	n_{s1}	$n_{s2} \dots$	n_{sk}	$\dots n_{sm}$	$n_{s.}$

¹Uma linha representa um indivíduo ou uma combinação de variáveis

² n_{jk} é o número de observações na k ésima categoria de resposta na j ésima linha;

$$^3n_{j.} = \sum_{k=1}^m n_{jk}$$

Os dados na tabela de contingência podem ser representado simbolicamente por uma matriz $m \times s$:

$$Y = |y_1, y_2 \dots y_j, \dots, y_s|, \text{ em que } y_j \text{ é um vetor } m \times 1.$$

Tem-se que:

$y_j = \sum_{r=1}^{n_j} y_{jr}$ e y_{jr} é um vetor $mx1$, tendo 1 na linha correspondente à categoria de resposta da $j^{\text{ésima}}$ unidade experimental e zero nos outros lugares.

Y tem distribuição conjunta com o vetor de parâmetros θ , sendo a densidade conjunta $f(Y, \theta)$. Pelo teorema de Bayes tem-se que: $f(\theta|Y) = g(Y|\theta) \cdot p(\theta)/t(Y)$, em que $t(Y)$ é a densidade marginal de Y ; $p(\theta)$ é a densidade “a priori” de θ , que reflete a relativa incerteza acerca de θ antes que os dados Y estejam disponíveis; $g(Y|\theta)$ é a função de verossimilhança e $f(\theta|Y)$ é a densidade “a posteriori”. Como $t(Y)$ não varia com θ , a densidade “a posteriori” pode ser usada como: $f(\theta|Y)g(Y|\theta) \cdot p(\theta)$

Associado com cada linha da tabela de contingência (sxm), existe um parâmetro de locação “ η_j ”, tal que a variável subjacente para a $j^{\text{ésima}}$ unidade experimental na $j^{\text{ésima}}$ linha possa ser escrita como:

$$\text{Predisposição} = \ell_{jq} = \eta_j + \varepsilon_{jq}, \text{ em que:}$$

$$j=1, \dots, s \text{ e } q=1, \dots, \eta_j \text{ e } \varepsilon_{jq} \sim N(0, I\sigma^2).$$

Nota-se que a variância residual é tomada como a unidade de medida na escala subjacente. Além disso, ao parâmetro de locação η_j é dada uma estrutura linear, como por exemplo a do modelo animal ou modelo de reprodutor.

Componentes de variância podem ser estimados usando-se um algoritmo tipo EM (Maximização das Esperanças) para aproximar a Máxima Verossimilhança Marginal (MML), considerada por Harville & Mee (1984) como uma extensão natural do método da Máxima Verossimilhança Restrita (REML). Para isso pode-se usar o “CMMAT2”, que é um programa para análises de modelos lineares mistos e de limiar, proposto por Misztal (1989), com suporte para estimação de componentes de variância “tipo REML” e modelo reprodutor avô-materno. Neste programa é empregado um algoritmo iterativo, implementado na linguagem de programação *Fortran 77*, fazendo uso do algoritmo EM.

2. Resultados de Pesquisas com Aplicações de Modelos Lineares e Não-Lineares, na Análise de Características Categóricas, em Melhoramento Animal

De acordo com Van Vleck & Edlin (1984), a maioria das avaliações de touros para dificuldade de parto tem sido realizada com base em escores definidos em uma escala linear; freqüentemente os escores estão no intervalo de 1 a 5, ou seja, parto fácil até a dificuldade extrema. As mensurações tem sido tratadas como uma simples variável contínua com escores discretos. Análises semelhantes, com base nos escores, têm sido feitas ajustando-se para os vários fatores identificados e que influenciam a dificuldade de nascimento, tais como: sexo de bezerro, idade da mãe ao parto (vaca ou novilha) e tamanho da mãe.

Com isso, esses autores estudaram como estender a análise de características múltiplas considerando-se dados categóricos para o caso em que cada categoria das combinações entre sexo do bezerro e idade da mãe ao parto é considerada como uma característica distinta. Por exemplo, com cinco categorias, dois sexos e três partos, o número total de características pode ser 30. As dificuldades em estimar as covariâncias entre as características é o motivo para tentar simplificar o método.

Com base nos resultados encontrados, os autores concluíram que os modelos de limiar apresentados por Harville e Mee (1984) e por Gianola e Foulley (1983) são mais apropriados, entretanto, as exigências computacionais requeridas pelos procedimentos destes autores são maiores.

Estudos com simulação e análises de dados reais (MEIJERING; GIANOLA, 1985; DJEMALI et al., 1987) tem sido feito, para testar as condições nas quais o modelo de limiar possa ser mais eficiente que o modelo linear. Por isso, os autores compararam a eficiência do modelo linear e de limiar, estimando-se os componentes de variância e os valores genéticos de touros da raça Charolês, utilizados em um programa de teste de progênie, para o efeito direto de distocia.

Os resultados dos partos foram classificados em quatro categorias: 1) sem ajuda; 2) com pouca ajuda; 3) com muita ajuda; 4) cesariana. Foram definidas, assim, duas variáveis: tipo de parto (1 a 4) e distocia, representada em dois grupos, de acordo com a facilidade (categorias 1 e 2) ou dificuldade de parto (categorias 3 e 4). Ambas as variáveis foram analisadas considerando-se os dois modelos, linear (LM) e de limiar (TM).

Utilizando-se um modelo linear (LM) o efeito de touro foi estimado empregando-se a metodologia BLUP (HENDERSON, 1973), após os componentes de variância terem sido estimados pelo método REML.

Com o modelo de limiar (TM) as avaliações dos touros foram obtidas pelo método GFCAT (GIANOLA; FOULLEY, 1983) e os componentes de variância foram estimados pelo método da Máxima Verossimilhança Marginal usando-se um algoritmo EM.

Os métodos foram comparados por meio das estimativas de herdabilidades e correlação de ordem entre as avaliações de touros. Para a variável distocia, as comparações foram realizadas, somente, utilizando-se todo o conjunto de dados disponível. Por outro lado, para a variável tipo de parto as comparações foram feitas considerando-se o arquivo completo e também quatro arquivos reduzidos, nos quais apenas 12% a 18%, aproximadamente, de progênies por touros, foram selecionadas aleatoriamente.

As estimativas de herdabilidades e os coeficientes de correlação de ordem entre as avaliações de touros por ambos os modelos, TM e LM, estão apresentados no quadro 2. Concordando com os resultados obtidos por Dempster e Lerner (1950), para características que apresentam apenas duas classes e, com os encontrados por Gianola (1982), para características com mais de duas classes, a herdabilidade estimada foi maior quando um modelo de limiar foi utilizando, em comparação com as estimativas obtidas por meio de um modelo linear.

Meijering e Gianola (1985) concluíram que o modelo de limiar não incrementou, significativamente, a eficiência na seleção de touros, utilizados em teste de progênie, na França, com inseminação artificial. Entretanto, observaram que, com menos informações, como em avaliações genéticas de vacas ou avaliando-se touros, quando os registros são provenientes de acasalamentos com monta natural, o modelo de limiar deve ser preferido em relação aos modelos lineares. Contudo, as exigências computacionais podem ser um importante obstáculo nas avaliações, utilizando-se modelos de limiar. Por isso, melhorias na área computacional são necessárias, para que as avaliações, em gado de corte, para disto-

cia, seja realizada por meio de um modelo de limiar.

Em produção de suínos, o tamanho de leitegada é

uma das características mais importantes. Nos programas de melhoramento, em vários países, o número de leitões nascidos vivos é a característica com o maior peso econômico

Quadro 2: Estimativas de herdabilidades e coeficientes de correlação de ordem, obtidos na avaliação de touros

	Arquivo completo		Média dos 4 arquivos reduzidos
	h ² para tipo de parto	h ² para distocia	h ² para tipo de parto
Modelo Linear (LM)	0,127	0,089	0,133
Modelo de Limiar (LM)	0,159	0,263	0,178
Correlação de ordem entre LM e TM	0,999	0,992	0,970

relativo e a avaliação genética para esta característica é, na maioria das vezes, realizada utilizando-se o BLUP.

Quando as variâncias são conhecidas, a seleção realizada com base nos valores genéticos estimados por meio do BLUP maximiza os ganhos genéticos esperados. O tamanho de leitegada em suínos, entretanto, é uma característica discreta e deve ser analisada como um variável categórica. Modelos não-lineares, como o modelo de limiar (TM) (GIANOLA; FOULLEY, 1983) e o modelo de Poisson (PM) tem sido propostos para estimação de valores genéticos para características categóricas.

Apesar do desenvolvimento das propriedades teóricas, os modelos não-lineares não têm sido implementados nas rotinas de avaliação genética em programas de melhoramento. A principal razão para a não utilização desses modelos é que são mais complicados e demandam mais recursos computacionais que os modelos lineares (LM). Entretanto, os recentes avanços dos recursos computacionais e o desenvolvimento de programas mais eficientes, podem fazer com que a aplicação de modelos TM e PM sejam amplamente difundidos em programas de melhoramento (MISZTAL et al., 1989).

Assim, Perez-Enciso et al. (1993) compararam modelos lineares mistos (LM) com um modelo de Poisson (PM), na avaliação genética de suínos, para número de leitões nascidos vivos e número total de nascimentos, utilizando-se registros de 8.056 leitegadas. Com base nos resultados, os autores concluíram que os modelos LM e PM tiveram performance semelhantes nas análises. Neste estudo, ao se utilizar o PM, foram encontradas dificuldades na estimação de componentes de variância. Provavelmente, devida à aproximação normal que é utilizada no algoritmo que implementa o método REML ou, ainda, pelo número limitado de informações utilizadas no estudo. Por esta razão, o PM parece não ser a melhor alternativa para análises genéticas de tamanho de leitegada.

Trabalhando também com suínos, Olesen et al. (1994), compararam modelos lineares mistos (LM), modelo de limiar (TM) e modelo de Poisson (PM), na avaliação genética de animais, para número de leitões nascidos. A herdabilidade estimada para esta característica foi de 0,20, 0,39 e 0,08, utilizando-se LM, TM e PM, respectivamente. A correlação de ordem de Spearman para os valores genéticos de reprodutores, entre LM e TM foi 0,998, indicando que as

classificações foram, praticamente, as mesmas.

Os resultados de PEREZ-ENCISO et al. (1993) estão de acordo com os encontrados em uma comparação envolvendo características categóricas em bovinos (JENSEN, 1986; DJEMALI et al., 1987). Meijering e Gianola (1985) mostraram, em um estudo de simulação, que a seleção com base nas avaliações genéticas realizadas utilizando-se TM, produziu maiores ganhos quando a seleção foi feita com base em um modelo linear, considerando-se uma característica binária e que os dados eram desbalanceados.

As correlações de ordem, para os valores genéticos, obtidas entre PM e LM e, entre PM e TM, foram de 0,939 e 0,927, respectivamente. Estas correlações, menores que as correlações entre LM e TM, indicam mudanças nas classificações, quando a análise é realizada utilizando-se PM.

Olesen et al. (1994) verificaram que os resultados obtidos, utilizando os modelos lineares e de limiar, foram semelhantes. O modelo de Poisson não apresentou resultados satisfatórios, nas análises genéticas, para tamanho de leitegada. O modelo de limiar proporcionou maiores estimativas de herdabilidade e aumento na acurácia das predições, comparadas com um modelo linear. Assim, os autores concluíram que o modelo de limiar pode ser utilizado em análises genéticas para tamanho de leitegada.

Utilizando registros de ovinos da raça Merino, na África do Sul, Olivier et al. (1998) empregaram um modelo de limiar para estimar herdabilidades e prever valores genéticos de reprodutores para características reprodutivas e de sobrevivência.

As estimativas de herdabilidades e a predição dos valores genéticos dos reprodutores foram obtidos por meio do programa GFCAT. O GFCAT é um conjunto de programas para análises de modelos de limiar mistos com suporte em estimação de componentes de variância do tipo REML, com base nos métodos de Gianola e Foulley (1983). Por este modelo, as respectivas características são o resultado de uma observação fenotípica subjacente, que excedem a um dado limiar. As características fenotípicas, contínuas e não-observadas são consideradas como tendo distribuição normal.

Por meio dos resultados encontrados, os autores concluíram que o modelo de limiar pode ser eficientemente utilizado para obter as estimativas dos parâmetros genéticos e para prever valores genéticos de reprodutores para características reprodutivas. Destacaram ainda que, quando a pre-

dição de valores genéticos por meio de um modelo animal de limiar estiver disponível, proporcionará um incremento na resposta à seleção, uma vez que todos os animais poderão ser selecionados com base nos seus valores genéticos, para a característica de interesse.

Uma vantagem adicional do uso dos valores genéticos, obtidos por meio de um modelo animal é que estes podem ser estimados mais cedo que os obtidos com um modelo de reprodutor. Além disso, a seleção de reprodutores pode ser feita com base na performance de sua mãe.

No Brasil, Souza (1998), utilizando registros de ovinos deslançados da raça Santa Inês, comparou estimativas de componentes de variância e parâmetros genéticos para características de reprodução, aplicando modelos lineares (modelos reprodutor e animal) e não-linear (modelo de limiar: modelo de reprodutor). Componentes de variância para os modelos lineares foram estimados usando-se o método REML, enquanto no modelo de limiar eles foram estimados usando a MML, por meio dos programas MTDFREML e CMMAT2, respectivamente. Os procedimentos utilizados nas análises foram apresentados por Gianola & Foulley (1983), para dados categóricos (modelo de limiar) e por Henderson (1963, 1973 e 1974), para modelos lineares, respectivamente.

As estimativas de herdabilidades obtidas com o modelo de limiar foram maiores do que aquelas obtidas com os modelos lineares para todas as características. As herdabilidades obtidas por meio do modelo linear de reprodutor para fertilidade ao parto, número de crias nascidas e número de crias desmamadas foram, respectivamente, 0,07, 0,07 e 0,04. Na mesma ordem, esses valores foram 0,03, 0,09 e 0,01, quando se utilizou o modelo animal. Correspondentemente, as herdabilidades estimadas com o modelo de limiar foram 0,12, 0,13 e 0,05. As estimativas de repetibilidades foram variadas e algumas delas foram menores do que as herdabilidades, principalmente quando foram utilizados os modelos de reprodutor (linear e de limiar). Correlação de ordem de Spearman, obtidas entre os valores genéticos para os reprodutores, com base nos modelos lineares e de limiar, foram altas e próximas da unidade, evidenciando semelhante classificação dos reprodutores. A despeito da similaridade entre os dois modelos (linear reprodutor e de limiar), o de limiar parece mais adequado em análises genéticas para fertilidade, número de crias nascidas e desmamadas, por produzir estimativas de herdabilidades maiores em até 80%, além de considerar a natureza binária das características. Adicionalmente, essas herdabilidades mais altas, obtidas na escala normal (modelo de limiar), implicariam em uma maior acurácia na predição.

Neste mesmo trabalho, Souza (1998), comparou aplicação de modelos lineares e de limiar em análises genéticas para sobrevivência de crias, com o objetivo de estimar componentes de variância e herdabilidade para sobrevivência da cria até o desmame, considerando a sobrevivência como uma característica da cria. Componentes de variância para os modelos lineares foram estimados usando-se o método REML, enquanto no modelo não linear (modelo de limiar) eles foram estimados usando-se uma aproximação da MML, por meio do programa CMMAT2. A estimativa de herdabilidade, estimada por meio do modelo de limiar (0,29), foi duas vezes maior do que aquela estimada com o modelo linear.

Correlação de ordem de Spearman entre os valores genéticos dos reprodutores, com base nos dois modelos foi 0,96, mostrou semelhante classificação dos reprodutores pelas duas metodologias. Considerando as estimativas obtidas neste estudo, com o modelo de limiar, verifica-se que é possível obter ganho genético para sobrevivência por meio da seleção.

Comentários

Com base na revisão de literatura realizada pode-se verificar que a utilização do modelo de limiar, que considera a natureza multinomial das características é adequado para utilização em avaliações genéticas de animais, para características categóricas.

Por meio do modelo de limiar obtém-se maiores estimativas de parâmetros genético que os obtidos pelos modelos lineares, proporcionado, assim, a obtenção de maiores ganhos genéticos por meio da seleção.

Além disso, os valores genéticos dos animais para características categóricas podem ser preditos com maior acurácia, quando se utiliza um modelo de limiar.

Referências

- BERGER, P. J.; FREEMAN, A. E. Prediction of sire merit for calving difficulty. *J. Dairy Sci.* v. 61, p. 1146-1150, 1978.
- CAÑÓN, J.; FUENTE, F. **Evaluacion de reproductores para caracteres discretos.** [S.l.: s.n.], [19--]. p.217-226.
- DEMPSTER, E. R.; LERNER, I. M. Heritability of threshold characters. *Genetics*, v. 35, p. 212-236, 1950.
- DJEMALI, M.; BERGER, P. J.; FREEMAN, A. E. Ordered categorical sire evaluations for dystocia in Holsteins. *J. Dairy Sci.* v. 70, p. 2374, 1987.
- FALCONER, D. S. The inheritance of liability to certain diseases estimated from the incidence from relatives. *Ann. Hum. Genet.* v. 29, p. 51-76, 1965.
- FALCONER, D. S. **Introdução a genética quantitativa.** Vinosa: UFV, 1987. 279 p.
- FOULLEY, J. L.; GIANOLA, D.; THOMPSON, R. Prediction of generatic merit from data on binary and quantitative variates na application to calving difficulty, borth weight and pelvic apening. *Génét. Sé. Evol.* v. 15, p. 401-424, 1983.
- FOULLEY, J. L.; GIANOLA, D. Estimation of genetic merit from bivariate "all or none" responses. *Génét. Sé. Evol.* v. 16, p. 285-306, 1984.
- GIANOLA, D. A method of sire evaluation for dichotomies. *J. Anim. Sci.* v. 51, p. 1266-1271, 1980a.
- GIANOLA, D. Genetic evaluation of animals for traits with categorical responses. *J. Anim. Sci.* v. 51, p. 1272-1276, 1980b.
- _____. Theory and analysis of threshold characters. *J. Anim.*

Sci. v. 54, p. 1079, 1982.

GIANOLA, D.; FOULLEY, J. L. Sire evaluation for ordered categorical data with a threshold model. **Génét. Sél. Evol.** v. 15, p. 201-224, 1983.

GIANOLA, D.; NORTON, H. W. Scaling threshold characters. **Genetics**, v. 99, p. 357-364, 1981.

GRIZZLE, J. R.; STARMER, C. F.; KOCH, G. C. Analysis of categorical data by linear models. **Biometrics**, v. 25, p. 489-504, 1969.

HARVILLE, D. A.; MEE, R. W. A mixed model procedure for analyzing ordered categorical data. **Biometrics**, v. 40, p. 393-408, 1984.

HENDERSON, C. R. Selection index and expected genetic advance. In: NATIONAL ACADEMY OF SCIENCE/ NATIONAL RESEARCH COUNCIL. **Statistical genetics and plant breeding**. Washington: NAS/NRC, 1963. p. 141-163.

_____. Sire evaluation and genetic trends. In: ANIMAL BREEDING GENETICS SYMPOSIUM IN HONOR OF Dr. J. L. LUSH, 1973, Blacksburg. **Proceedings...** Champaign: ASAS/ ADSA, 1973. p. 10-41.

_____. General flexibility of linear model techniques for sire evaluation. **J. Dairy Sci.** Champaign, v. 57, p. 963-972, 1974.

HOESCHELE, I.; TIER, B.; GRASER, H-U. Multiple trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. **J. Anim Sci.** v. 73, p. 1609-1627, 1995.

JENSEN, J. Sire evaluation for type traits with linear and non-linear procedures. **Livestock Production Science**, v. 15, p. 165-171, 1986.

KOCH, G. G.; REINFURT, D. W. The analysis of categorical data from mixed models. **Biometrics**, v. 27, p. 157, 1971.

MEIJERING, A.; GIANOLA, D. Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. **Génét. Sél. Evol.** v. 17, p. 115-132, 1985.

MISZTAL, I.; GIANOLA, D.; FOULLEY, J. L. Computing aspects of nonlinear method of sire evaluation for categorical data. **J. Dairy Sci.** v. 53, p. 643-650, 1989.

OLESEN, I. et al. A comparison of animal and nonnormal mixed models for number of lambs born in Norwegian sheep. **J. Anim. Sci.** v. 72, p. 1166-1173, 1994.

OLIVIER, W. J. et al. Genetic parameter estimates for fitness traits in South African Merino Sheep. **Livestock Production Science**, v. 56, p. 71-77, 1998.

PEREZ-ENCISO, M.; TEMPELMAN, R. J.; GIANOLA, D. A comparison between linear and Poisson mixed models for litter size in iberian pigs. **Livestock Production Science**, v. 35, p.

303-316, 1993.

POLLAK, E. J.; FREEMAN, A. E. Parameter estimation and sire evaluation for dystocia and calf size in Holsteins. **J. Dairy Sci.** v. 59, p. 1817-1824, 1976.

QUAAS, R. L.; VAN VLECK, L. D. Categorical traits sire evaluation by best linear unbiased prediction of future progeny category frequencies. **Biometrics**, v. 36, p.117-122, 1980.

SCHAEFFER, L. R.; BURNSIDE, E. B. Survival rates of tested daughters of sires in artificial insemination. **J. Dairy Sci.** v. 59, p. 554, 1974.

SCHAEFFER, L. R.; WILTON, J. W. 1976. Methods of sire evaluation for calving ease. **J. Dairy Sci.** v. 59, p. 544-551, 1976.

SCHAEFFER, L. R. et al. Evaluation of holstein-friesian dairy sires for conformation of their daughters. **Can. J. Anim. Sci.** v. 58, p. 409-417, 1978.

SNELL, E. J. A scaling procedure for ordered categorical data. **Biometrics**, v. 20, p. 592-607, 1964.

SOUZA, W. H. **Aplicação de modelos lineares e não-lineares em características de reprodução, sobrevivência e crescimento em ovinos da raça Santa Inês**. 1997. 131 f. Tese (Doutorado) - Universidade Federal de Minas Gerais, Belo Horizonte, 1997.

THOMPSON, R. Sire evaluation. **Biometrics**, v. 35, p. 339-353, 1979.

TONG, A. K.W.; WILTON, J. W.; SCHAEFFER, L. R. Evaluation of calving ease for Charolais sires. **Can. J. Anim. Sci.** v. 56, p. 12-26, 1976.

VAN VLECK, L. D.; EDLIN, K. M. Multiple trait evaluation of bulls for calving ease. **J. Dairy Sci.** v. 67, p. 3025-3033, 1984.

VAN VLECK, L. D.; KARNER, P. J. Sire evaluation by BLUP for categorically scored type traits. **J. Dairy Sci.** v. 63, p. 1328-2341, 1980.

WRIGHT, S. An analysis of variability in number of digits in an inbred strain of guinea pigs. **Genet.** v. 19, p. 506-536, 1934.

Recebido em: 30/06/2006

Aceito em: 20/11/2007