

## AGRÁRIA

Revista Brasileira de Ciências Agrárias

v.3, n.4, p.381-385, out.-dez., 2008

Recife, PE, UFRPE. www.agraria.ufrpe.br

Protocolo 366 - 02/04/2008 • Aprovado em 13/09/2008

Elmer F.V. Tapia<sup>2</sup>

A. Ramos<sup>2</sup>

Francisco S. Weschler<sup>3</sup>

Alencariano J. S. Falcão<sup>2</sup>

Danísio P. Munari<sup>4</sup>

# Detecção e caracterização da heterogeneidade de variância na produção de leite de bovinos da Raça Holandesa<sup>1</sup>

## RESUMO

Este estudo propõe avaliar e caracterizar a heterogeneidade de variância na produção de leite de bovinos da raça Holandesa em rebanhos estratificados segundo o nível de produção. Utilizaram-se 12.416 registros de lactações distribuídas em 92 rebanhos dos estados de São Paulo e Paraná. Os dados originais e transformados pela raiz quadrada e logarítmica foram analisados depois de ajustados para os efeitos de rebanho, ano-estação de parição, duração de lactação e efeito linear e quadrático da idade da vaca ao parto. Na escala original os resultados evidenciaram heterogeneidade de variâncias do tipo regular e irregular entre os três níveis de produção, as quais tendem a aumentar linearmente conforme aumenta o nível de produção. A transformação raiz quadrada foi efetiva para atenuar a heterocedasticidade do tipo regular mas não a do tipo irregular, porém a transformação logarítmica não foi adequada para atenuar a heterogeneidade de variâncias.

**Palavras-chave:** bovinos de leite, heterocedasticidade, modelos

## Detection and characterization of the heterogeneity of variance in milk production of Holstein cows

## ABSTRACT

The objective of this study was to evaluate and to characterize the variance heterogeneity in the milk production of Holstein cows, in herds stratified according to the production level. 12,416 records of 92 herds belonging to the States of São Paulo and Paraná were analyzed. The original data and transformed by the square and logarithmic root, and were analyzed after being adjusted for the herds effects, year-station parity, lactation duration and linear and quadratic effect of the age of the cow to the parity. In the original scale the results evidenced heterogeneity of variances of the regular and irregular type among the three production levels and a linear association between the variances a the means. The square root transformation was then chosen to attenuate the heterocedasticity. This transformation was effective to attenuate the heterocedasticity of the regular, but the logarithmic transformation was not.

**Key words:** dairy cattle, heterocedasticity, models

<sup>2</sup> Universidade Federal do Tocantins - Escola de Medicina Veterinária e Zootecnia, BR 153, Km 112, CP 132, Araguaína, TO. eltavale@hotmail.com

<sup>3</sup> UNESP - Depto de Produção e Exploração Animal, Distrito Rubião Jr s/n, Botucatu, SP, 18.618-000

<sup>4</sup> UNESP - Depto de Ciências Exatas, Via de Acesso Prof. Paulo Donato Castellane s/n, Jaboticabal, SP

<sup>1</sup> Parte da Tese de Doutorado do primeiro autor apresentada ao Instituto de Biotecnologia da UNESP-Botucatu

## INTRODUÇÃO

Os modelos de avaliação genética em populações de bovinos se tem tornado cada vez mais complexos. A maioria dos programas de melhoramento utiliza a metodologia dos modelos mistos de Henderson para avaliar o mérito genético dos candidatos à seleção. Até recentemente, grande parte das avaliações partia da premissa de que as variâncias entre as subclasses eram constantes, mas vários estudos ressaltam a existência de heterogeneidade de variância genética, residual, de ambiente permanente e fenotípica para características produtivas e de conformação, em relação à região, nível de produção, grupo de contemporâneos e rebanho, dentre outros fatores (Famula, 1989; Boldman & Freeman, 1990; Falcão, 2006).

Quando a heterocedasticidade é do tipo regular e se conhece a distribuição dos dados, pode-se recorrer a transformações a fim de corrigir a heterogeneidade de variâncias; assim, por exemplo, se o desvio padrão varia diretamente com a média de produção do rebanho, a transformação logarítmica é adequada para estabilizar as variâncias. No caso de uma relação linear entre as variâncias e as médias, a transformação raiz quadrada seria a recomendada (Bartlett, 1936; Gill, 1981; Hill, 1984; Torres et al., 1999).

Existem inúmeras alternativas para detectar a existência de heterogeneidade de variância, dentre as quais se pode destacar: a 1) inspeção dos gráficos dos desvios padrões em função das médias (Kung, 1988) ou das subclasses rebanho-ano (van der Werf et al., 1994); 2) aplicação de testes, tais como os testes de Hartley, Bartlett, Levene e Box (Milliken & Johnson, 1984) e 3) por meio da estratificação dos dados. Se os rebanhos são os fatores que causam a heterogeneidade de variâncias, esta pode ser verificada facilmente após estratificar os rebanhos em diferentes níveis de produção (Famula, 1989; Boldman & Freeman, 1990; Tapia et al., 1996; Costa et al., 1999). A estratificação pode basear-se nas médias ou variâncias fenotípicas das subclasses rebanho, rebanho-ano ou rebanho-ano-estação.

Objetivou-se, neste estudo, avaliar e caracterizar a heterogeneidade de variâncias da produção de leite de bovinos da raça Holandesa.

## MATERIAL E MÉTODOS

Analisaram-se 12.416 registros de lactação de 3.632 vacas puras de origem da raça Holandesa, que pariram no período de 1976 a 1987, filhas de 244 touros, distribuídas em 92 rebanhos nos estados de São Paulo e Paraná, cujos dados foram cedidos pelo Serviço de Controle Leiteiro da Associação Brasileira de Criadores de Bovinos da Raça Holandesa.

A produção de leite na escala original, transformada por raiz quadrada e logarítmica, foi ajustada para os fatores ambientais por meio do modelo I,

$$y_{ijkl} = \mu + R_i + AE_j + b_1(O_{ijk} - \bar{O}) + b_2(O_{ijk} - \bar{O})^2 + b_3(D_{ijk} - \bar{D}) + \varepsilon_{ijkl} \quad (1)$$

em que  $y_{ijkl}$  é produção de leite da vaca (PL)  $l$ ;  $\mu$  = média geral;  $R_i$  = efeito fixo do  $i$ -ésimo rebanho;  $AE_j$  = efeito fixo da  $j$ -ésima subclasse ano-estação de parição;  $b_1$  e  $b_2$  são os coeficientes de regressão linear e quadrática da produção de leite em função da idade da vaca ao parto;  $b_3$  = coeficiente de regressão linear da produção de leite em função da duração da lactação;  $O_{ijk}$  = efeito da  $k$ -ésima idade da vaca ao parto;  $D_{ijk}$  = efeito da  $k$ -ésima duração da lactação;  $\bar{O}$  = média das idades do parto;  $\bar{D}$  = média da duração de lactação e  $\hat{\varepsilon}_{ijkl}$  = erro aleatório associado à observação  $Y_{ijkl}$ . Como complemento desta análise se obtiveram o valor predito de cada lactação e as médias ajustadas dos rebanhos nas três escalas.

As médias ajustadas da PL dos rebanhos foram utilizadas para classificá-los em três níveis de produção, a saber: baixo, médio e alto, por meio da análise de agrupamento do SAS<sup>®</sup> (SAS, 1993), utilizando o modelo II,

$$M_{ij} = \mu + R_i + \varepsilon_{ij} \quad (2)$$

em que  $M_{ij}$  é a média ajustada da PL por rebanho;  $\mu$  = média geral;  $R_i$  = efeito fixo do  $i$ -ésimo rebanho e  $e_{ij}$  = desvio aleatório associado à  $M_{ij}$  informação.

A avaliação e a caracterização da heterocedasticidade dos dados na escala original, transformados pela raiz quadrada e logarítmica, foram realizadas por meio da análise gráfica das variâncias em função das médias ajustadas dos rebanhos e pelo teste de Levene, aplicando-se o seguinte modelo III,

$$Z_{ijk} = \mu + N_i + R_{ij} + \varepsilon_{ijk} \quad (3)$$

em que  $Z_{ijk}$  é o desvio absoluto da produção de leite observada em relação ao valor predito;  $\mu$  = efeito da média geral;  $N_i$  = efeito fixo do  $i$ -ésimo nível de produção;  $R_{ij}$  = efeito fixo do  $j$ -ésimo rebanho dentro do  $i$ -ésimo nível de produção e  $e_{ijk}$  = efeito do erro associado à informação  $Z_{ijk}$ .

Avaliou-se a transformação mais adequada para atenuar a heterocedasticidade dos dados na escala original por meio do modelo de regressão IV,

$$S_{ij} = b_0 + b_1 M_i + b_2 M_i^2 + \varepsilon_{ij} \quad (4)$$

em que  $S_{ij}$  = variância do  $i$ -ésimo rebanho;  $b_0$  = intercepto;  $b_1$  e  $b_2$  são os coeficientes de regressão linear e quadrático;  $M_i$  = média ajustada do  $i$ -ésimo rebanho e  $e_{ij}$  = erro.

## RESULTADOS E DISCUSSÃO

Na Tabela 1 se apresenta a análise de variância do efeito dos fatores ambientais na produção de leite dos dados originais, cujos resultados indicam que todas as variáveis incluídas no modelo exercem efeito significativo ( $p < 0,01$ ) sobre a produção de leite, concordando com as pesquisas realizadas por Conceição Junior et al. (1993), Almeida et al. (1995) e Ribas et al. (1996). O valor do coeficiente de determinação mostra que as variáveis incluídas no modelo explicam 55% da

**Tabela 1.** Análise de variância da produção de leite na escala original

**Table 1.** Analysis of variance of milk production in original scale

Fonte de variação	GL	Quadrado médio
Rebanho	91	75.843.464**
Ano-estação	47	17.250.325**
Idade ao parto		
Linear	1	1.861.395.783**
Quadrática	1	711.865.633**
Duração da lactação		
Linear	1	9.443.460.382**
Resíduo	12.274	1.383.751
R <sup>2</sup> , %		55

\*\*p < 0,0001

variação total da produção de leite, valores esses considerados adequados para este tipo de característica.

Apresenta-se, na Tabela 2, as médias ajustadas para os fatores ambientais, o número de lactações e as variâncias dos três níveis de produção. O maior número de rebanhos e lactações no nível médio de produção era previsto, uma vez que, a produção de leite, tal como outras características quantitativas de interesse zootécnico, tendem à distribuição.

**Tabela 2.** Variância, média, número de lactações e rebanhos de acordo com a produção de leite, na escala original

**Table 2.** Variance, mean, number of records and herds, according to milk yield in original scale

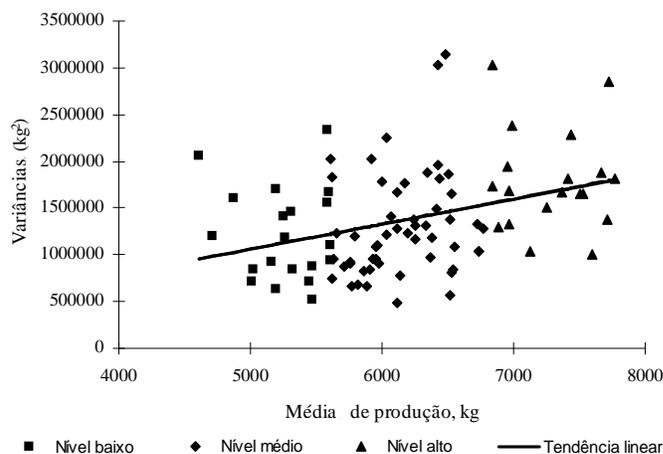
Nível de produção	Número de rebanhos	Número de lactações	Média kg	Variância kg <sup>2</sup>
Baixo	20	2.441	5.312	1.975.205
Médio	53	7.358	6.325	2.529.213
Alto	19	2.617	7.385	3.361.869
Total	92	12.416	6.296	3.038.049

A análise da Tabela 2 também evidencia uma associação positiva entre as médias de produção com as respectivas variâncias dos três estratos. De acordo com Gill (1981) e Falconer (1990), em muitas situações do campo biológico, a variância é proporcional à média de produção, razão por que se justifica o incremento da variância dos rebanhos com o aumento do nível de produção do leite.

A análise gráfica da heterocedasticidade dos dados originais entre os três níveis de produção confirma a heterogeneidade de variâncias entre os três níveis de produção e a associação positiva entre as médias e as variâncias dos estratos, como se observa na Figura 1, na qual também é possível verificar a coexistência dos dois tipos de heterocedasticidade, isto é, a do tipo regular e a irregular.

A avaliação da heterocedasticidade pelo teste de Levene dos dados originais é apresentada no Tabela 3, em que os resultados também confirmam a heterocedasticidade entre os três níveis de produção (p < 0,01) e entre os rebanhos de cada nível de produção (p < 0,01) observada na Figura 1; desta forma, para tornar independentes as variâncias e as médias houve necessidade de se aplicar determinada transformação.

A análise de regressão da Tabela 4, revela que a associação entre as variâncias e as médias é linear (p < 0,01); por



**Figura 1.** Análise gráfica da heterocedasticidade entre os três níveis de produção de leite na escala não transformada

**Figure 1.** Graphic analysis of heterocedasticity between milk yield levels in original scale

**Tabela 3.** Análise da heterocedasticidade entre e dentro dos níveis de produção pelo teste de Levene, na escala não transformada

**Table 3.** Analysis of heterocedasticity between and within milk yield level, using Levene test in original scale

Fonte de Variação	Grau de liberdade	Quadrado médio
Nível de produção	2	28.117.486**
Rebanho: Nível de produção	89	4.018.418**
Resíduo	12.324	507.071
R <sup>2</sup> , %		6,21

\*\*p < 0,01

**Tabela 4.** Análise da regressão das variâncias em função das médias dos rebanhos, na escala não transformada

**Table 4.** Regression analysis of variances according to herd yield milk mean in original scale

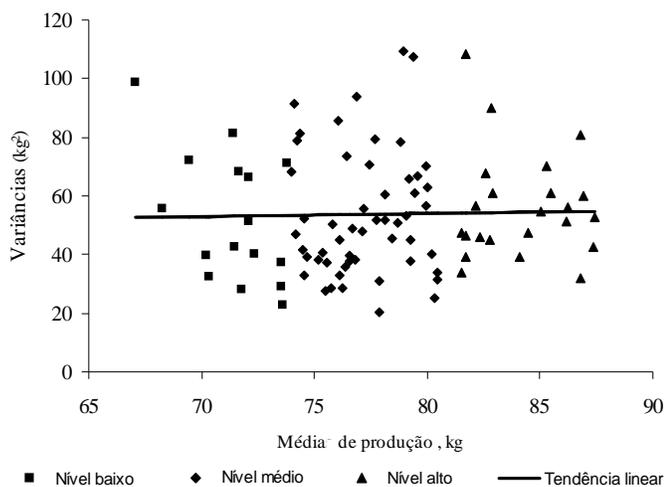
Fonte de Variação	Grau de liberdade	Quadrado médio
Linear	1	3.593.742.541.990**
Quadrático	1	439.831.567.495
Resíduo	89	290.917.230.633
R <sup>2</sup> , %		13

esta razão, a transformação pela raiz quadrada seria recomendada para atenuar a heterocedasticidade dos dados originais e não a transformação logarítmica, conforme recomendam Bartlett (1936) e Gill (1981).

Vários trabalhos mostraram as conseqüências da associação positiva entre as variâncias e as médias de produção dos rebanhos. Hill (1984) apontou que, ao se desconsiderar a heterogeneidade de variâncias entre os rebanhos, corre-se o risco de selecionar uma percentagem maior de animais naqueles rebanhos que têm maiores variâncias. O efeito se acentua conforme aumenta a pressão de seleção. Segundo Hill (1984) e Vinson (1987), a heterogeneidade de variâncias pode reduzir a precisão ou acurácia dos valores genéticos dos animais candidatos à seleção. Winkelman & Schaeffer (1987) mostraram que se as variâncias aumentam com a média de produção dos rebanhos e os touros com grande percentagem de suas

filhas em rebanhos de maior variabilidade podem ter seus valores genéticos superestimados.

Conforme se observa na Figura 2 e na Tabela 5, a transformação raiz quadrada tornou homogêneas as variâncias entre os três níveis de produção ( $p > 0,05$ ), pois a relação linear e positiva que existia entre as médias e as variâncias dos dados originais foi eliminada; assim, esta transformação tornou independentes as variâncias das médias. Ainda na Figura 2 e na Tabela 5 constata-se que, embora a transformação raiz quadrada tenha atenuado de maneira efetiva a heterocedasticidade do tipo regular não o fez com a heterocedasticidade do tipo irregular ( $p < 0,01$ ). Embora este resultado seja justificado pois, segundo Cochran (1947) as transformações surtem pouco ou nenhum efeito na heterocedasticidade do tipo irregular.



**Figura 2.** Análise gráfica da heterocedasticidade entre os três níveis de produção de leite na escala raiz quadrada

**Figure 2.** Graphic analysis of heterocedasticity between milk yield levels in root square scale

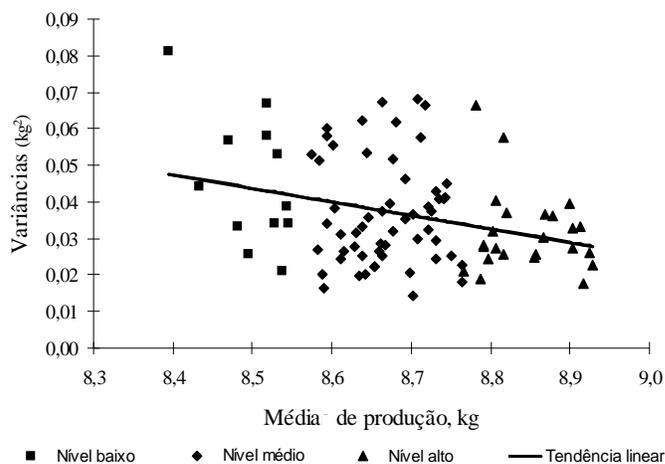
**Tabela 5.** Análise da heterocedasticidade entre e dentro os níveis de produção pelo teste de Levene, na escala raiz quadrada

**Table 5.** Analysis of heterocedasticity between and within milk yield level, using Levene test in root square scale

Fonte de variação	GL	Quadrado médio
Nível de produção	2	52,08
Rebanho: Nível de produção	89	158,98
Resíduo	12.324	19,31
R <sup>2</sup> , %		5,7

A Figura 3 indica que a transformação logarítmica não homogeneizou as variâncias entre os três níveis de produção; este resultado é justificado porque a transformação não foi recomendada para atenuar a heterocedasticidade. Boldman & Freeman (1990) e Torres et al. (1999) também não foram bem sucedidos ao aplicarem esta transformação.

Os resultados da heterocedasticidade pelo teste de Levene na escala logarítmica (Tabela 6) confirmam os resultados da Figura 3, haja vista que as variâncias tendem a decrescer com o nível de produção ( $p < 0,01$ ), da mesma forma como a



**Figura 3.** Análise gráfica da Heterocedasticidade entre os três níveis de produção de leite, na escala logarítmica

**Figure 3.** Graphic analysis of heterocedasticity between milk yield levels in logarithmic scale

**Tabela 6.** Análise da heterocedasticidade pelo teste de Levene, na escala logarítmica

**Table 6.** Analysis of heterocedasticity using Levene test in logarithmic scale

Fonte de variação	GL	Quadrado médio
Nível de produção	2	0,466
Rebanho: Nível de produção	89	0,109
Resíduo	12.324	0,0133
R <sup>2</sup> , %		6,1

heterogeneidade das variâncias entre os rebanhos de cada nível ( $p < 0,01$ ).

## CONCLUSÕES

Nos dados originais da produção de leite dos rebanhos leiteiros dos estados de São Paulo e Paraná constatou-se heterogeneidade de variâncias do tipo regular e irregular entre os três níveis de produção.

As variâncias tendem a se incrementar linearmente conforme aumenta o nível de produção dos rebanhos.

A transformação raiz quadrada foi efetiva para atenuar a heterocedasticidade do tipo regular mas não o foi a transformação logarítmica.

Em decorrência da associação positiva entre as médias e variâncias, sugerem-se estudos adicionais do efeito da heterogeneidade de variâncias no mérito genético dos animais, na precisão ou acurácia e nas taxas de ganho genético por ano.

## AGRADECIMENTOS

À Associação Brasileira de Criadores de Bovinos da Raça Holandesa, pela cessão dos dados, e aos revisores ad hoc, pelas valiosas sugestões apresentadas em prol da redação final do trabalho.

## LITERATURA CITADA

- Almeida, R.; Ribas, N. P.; Monardes, H. G. Estudo de características produtivas em rebanhos holandeses em primeira cria na região da Batavo, Paraná. In: Reunião da Sociedade Brasileira de Zootecnia, 32, 1995, Brasília. Anais...p.692-694.
- Bartlett, M. S. The square root-transformation in the analysis of variance. *Journal of the Royal Statistical Society*, v.3, n.1, p.68-78, 1936.
- Boldman, K. G.; Freeman, A. E. Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *Journal of Dairy Science*, v.73, p.503-512, 1990.
- Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, v.3, p.2-38, 1947.
- Conceição Junior, V.; Silva, H. M.; Pereira, C. S. Fatores ambientais e genéticos que afetam a produção de leite e de gordura em vacas da raça Holandesa. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v.45, n.1, p.81-95, 1993.
- Costa, C. N. An investigation into heterogeneity of variance for milk and fat yields of Holstein cows in Brazilian herd environments. *Genetics and Molecular Biology*, v. 22, n. 3, p. 375-381, 1999.
- de Veer, J. C.; van Vleck, L. D. Genetic parameters for first lactation milk yields at three levels of herd production. *Journal of Dairy Science*, v.70, p.1434-1441, 1987.
- Falcão, A. J. S.; Martins, E. N.; Costa, C. N.; Sakaguti, E. S.; Mazucheli, J. Heterocedasticidade entre estados para produção de leite em vacas da raça Holandesa, usando métodos bayesianos via amostrador de Gibbs. *Revista Brasileira de Zootecnia*, v.35, n.2, p.405-414, 2006.
- Falconer, D.S. Introduction to quantitative genetics. New York: Longman, 1990. 438p.
- Famula, T. R. Detection of heterogeneous variance in herd production groups. *Journal of Dairy Science*, v.72, p.715-721, 1989.
- Gill, J. L. Design and analysis of experiments in the animal and medical sciences. Ames: The Iowa State University Press, 1981. 410p.
- Hill, W. G. On selection among group with heterogeneous variance. *Animal Production*, v.39, n.3, p.473-477, 1984.
- Kung, F. H. Application of data transformation in forest genetics. *Silvae Genetica*, v.37, n.2, p.45-49, 1988.
- Milliken, G. E.; Johnson, D. E. Analysis of messy data, designed experiments. New York: van Nostrand Reinhold, 1984. V.1, 473p.
- Ribas, N. P.; Monardes, H.; Molento, C.; Almeida, R. Estudo dos efeitos de meio ambiente sobre as características produtivas de vacas da raça Holandesa no Estado de Paraná. In: Reunião da Sociedade Brasileira de Zootecnia, 23, 1996, Fortaleza. Anais...p.9-11.
- SAS Institute Inc. SAS User's Guide, Version 6, 2.ed., Cary, NC: SAS Institute Inc. 1993.
- Tapia, E. F. V.; Ramos, A. A.; Weschler, F. S.; Gonçalves, H. C. Heterogeneidade de variância das características de produção de bovinos da raça Holandesa no trópico. II. Estratificação por nível de produção das vacas. In: Reunião da Sociedade Brasileira de Zootecnia, 23, 1996, Fortaleza. Anais... 1996. p.68-70.
- Torres, R.A.; Bergmann, J.A.G; Costa, C. N.; Pereira, C. S; Valente, J; Penna, V. M.; Torres Filho, R A. Ajustamento para heterogeneidade de variância para a produção de leite entre rebanhos da raça Holandesa no Brasil. *Revista Brasileira de Zootecnia*, v. 28, n. 2, p. 295-303, 1999.
- van der Werf, J. H. J.; Meuwissen, T. H. E.; Jong, G. Effects of correction for heterogeneity of variance on bias and accuracy of breeding value estimation for Dutch dairy cattle. *Journal of Dairy Science*, v.77, p.3174-3184, 1994.
- Vinson, W. E. Potential bias in genetic evaluation from differences in variation within herds. *Journal of Dairy Science*, v.70, p.2450-2455. 1987.
- Winkelman, A.; Schaeffer, L. R. Effect of heterogeneity of variance on dairy sire evaluation. *Journal of Dairy Science*, v.71, p.3033-3039, 1988.