

Heterogeneidade de variância da produção de leite no dia do controle por agrupamento *k-means* utilizando modelos de regressão aleatória em vacas Gir

Donicer Montes Vergara^{1*}, Daniel Jordan de Abreu Santos¹, Daiane Cristina Becker Scalez², Naudin Hurtado-Lugo², Humberto Tonhati³, Aníbal Eugênio Vercesi Filho⁴

¹Pós-Graduação em Genética e Melhoramento Animal – FCAV/UNESP, Jaboticabal-SP. *U de Sucre. e-mail: donicermontes@hotmail.com

²Pós-Graduação em Zootecnia – FCAV/UNESP, Jaboticabal-SP

³Departamento de Zootecnia – FCAV/UNESP, Jaboticabal-SP

⁴Agência Paulista de Tecnologia dos Agronegócios – PRDTA/Nordeste Paulista, Mococa-SP

Resumo: O objetivo do presente trabalho foi utilizar o agrupamento *k-means* e descrever a curva de lactação de vacas Gir de alta e baixa produção utilizando modelos de regressão aleatória. Foram utilizados 38.220 registros de produção de leite no dia do controle (PLDC) das primeiras lactações de vacas Gir, provenientes do Programa Nacional de Melhoramento do Gir Leiteiro. Foi utilizado o modelo animal bicaracterístico de regressão aleatória, empregando-se polinômios de Legendre. A obtenção dos componentes de (co)variância foi realizada por inferência bayesiana. As estimativas de herdabilidade para o ambiente de alta e baixa produção descreveram comportamentos diferentes. O ambiente de alta produção apresentou valores baixos de variância fenotípica ao início da lactação, diferente do observado no ambiente de baixa produção. As estimativas de variância genética aditiva variaram no ambiente de alta produção, apresentando-se as maiores estimativas entre as primeiras PLDC. As variâncias residuais foram baixas no início da lactação, em ambos os ambientes. As correlações genéticas entre as produções de leite entre controles foram menores conforme aumentaram os controles, para os dois ambientes.

Palavras chave: gado de leite, inferência bayesiana, parâmetros genéticos, zebuínuo.

Heterogeneity of variance for test day milk yield by clustering *k-means* using random regression models in Gir cows

Abstract: The objective of this work was to use the *k-means* clustering and describe the curve of lactation Gir cows of high and low production using random regression models. Were used 38,220 records for test day milk yield of first lactations of Gir cows, from the National Program for Improvement of Gir Dairy. A two-trait animal model of random regression was used employing the Legendre polynomials. The obtaining of the components of (co)variance was performed by bayesian inference. Estimated heritabilities for the high and low production environment described different behaviors. The production environment high showed low values of phenotypic variance at the beginning of lactation, different from that observed in low production. The estimates of additive genetic variance ranged in high production, estimating the largest among test day milk yield. The residual variances were low at the start of lactation, in both environments.

Keywords: bayesian inference, dairy cattle, genetic parameters, zebuine.

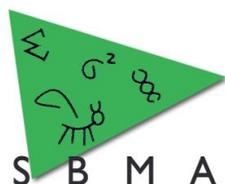
Introdução

As análises de agrupamento têm sido usadas para explorar a estrutura de conjuntos de dados e definir grupos homogêneos de animais dentro de uma população ou ambiente diferenciado. Entre os métodos de aglomeração estão os não hierárquicos, que se caracterizam por procurar maximizar a homogeneidade intragrupo, sem considerar a hierarquia entre grupos. As escolhas de um método de aglomeração e de uma medida de distância estão inter-relacionadas (Vicini & Souza, 2005). O objetivo do presente trabalho foi utilizar o agrupamento *k-means* e descrever a curva de lactação de vacas Gir de alta e baixa produção, utilizando modelos de regressão aleatória.

Material e Métodos

Foram analisados 38.220 registros de produção de leite no dia do controle (PLDC) das primeiras lactações de 4.994 vacas da raça Gir, provenientes do Programa Nacional de Melhoramento do Gir Leiteiro. Foram considerados os controles leiteiros entre o 6^o e 305^o dia de lactação. As PLDC foram divididas em 10 classes mensais. Os grupos de contemporâneos foram definidos por rebanho, ano e mês do controle.

Para discriminação em ambientes de alta e baixa produção, foi empregada a técnica de agrupamento de *k-means*. O número de médias foi $k=2$ e os centros iniciais dos *clusters* foram obtidos ordenando as distâncias e escolhendo as observações com intervalos constantes.



Os testes de Tukey para cada PLDC foram realizados para comparação entre os grupos, assim como os centróides destes, pelo teste de Hotelling. Foram utilizados modelos animais uni e bicaracterístico de regressão aleatória, empregando-se polinômios de Legendre. Os efeitos fixos foram compostos pelo grupo de contemporâneos, além de efeitos linear e quadrático da covariável idade da vaca (meses), aninhado em mês de controle e a curva média de lactação da população (modelada por um polinômio de Legendre de quarta ordem). Como efeitos aleatórios foram considerados o efeito genético aditivo, de ambiente permanente, além do residual. Para modelagem dos efeitos genéticos aditivos e de ambiente permanente foram utilizados polinômios de ordem quártica e quártupla, respectivamente. A variância residual foi modelada considerando-se a estrutura heterogênea, sendo consideradas seis classes (1, 2, 3-4, 5, 6, 7-10).

A obtenção dos componentes de (co)variância foi realizada por inferência bayesiana, empregando o programa GIBBS3F90 (Misztal, 2010). As estimativas de variâncias e herdabilidades para cada PLDC e as correlações entre elas foram obtidas por funções de covariâncias estimadas em cada ciclo, sendo possível obtenção de uma distribuição *a posteriori* para cada um destes parâmetros. Pelo teste de Heidelberg e Welch, foi determinado para cada parâmetro um período de burn-in, utilizando-se das amostras obtidas com o GIBBS3F90.

Resultados e Discussão

A representação da curva média de lactação foi similar à reportada por Herrera et al. (2008) para a raça Gir (Figura 1), sendo possível observar que o agrupamento proporcionou diferenças significativas, pelo teste de Tukey ($p < 0,05$), entre as PLDC. Pelo teste de Hotelling, também foi possível observar as diferenças entre os centróides dos grupos discriminados ($p < 0,05$). Resultados semelhantes foram relatados por Savegnago et al. (2012) pela análise de agrupamento não hierárquica.

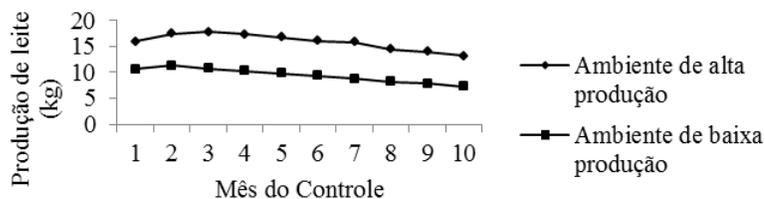


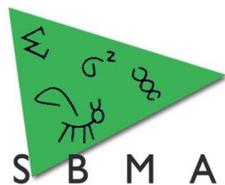
Figura 1. Curvas de lactação médias para dois ambientes de produção, discriminados pelo método de *k-means*.

As estimativas de herdabilidade para os ambientes de alta e baixa produção descreveram comportamentos diferentes durante a maior parte da lactação (Figura 2). Indicando que os intervalos de alta densidade apresentaram valores amplos para ambos os ambientes, especialmente até a quinta PLDC. A amplitude das estimativas de herdabilidade possivelmente esteve associada ao nível produtivo, magnitude e heterogeneidade de variância própria de cada ambiente.

Tendências semelhantes de herdabilidade no ambiente de baixa produção foram observadas por Herrera et al. (2008), no entanto, estes autores não utilizaram heterogeneidade de variância nos modelos de estudo.

As variâncias fenotípicas no ambiente de alta produção apresentaram valores baixos no início da lactação, tornando-se maiores na segunda metade da lactação, diferentemente do que ocorreu no ambiente de baixa produção, onde variaram menos no decorrer da lactação. Bignardi et al. (2009) utilizando um modelo que não utilizou heterogeneidade de variâncias nas análises, reportaram em bovinos da raça Holandesa nas primeiras lactações, que o comportamento das variâncias fenotípicas atingiu seus máximos valores ao início e ao final da lactação, resultados que diferem deste estudo. As estimativas de variâncias genéticas aditivas no ambiente de produção alta foram maiores entre as primeiras PLDC (resultados não apresentados). As estimativas de variâncias residuais foram baixas ao início da lactação, em ambos os ambientes, mantendo um padrão similar ao longo da lactação, esta tendência pode ser atribuída ao tipo de polinômio empregado (resultados não apresentados).

As estimativas de variâncias ambientais no ambiente de alta produção apresentaram tendência crescente até o quinto controle, retomando esse mesmo comportamento a partir do oitavo, até o final da lactação. O ambiente de baixa produção apresentou variâncias ambientais menores que as estimadas para o ambiente de alta produção, apresentando tendência ascendente a partir do sexto controle, exceto para a produção do



primeiro controle. Resultados contrários aos observados foram reportados por Herrera et al. (2008), que utilizaram modelos de regressão aleatória e duas funções paramétricas para o ajuste da curva de lactação.

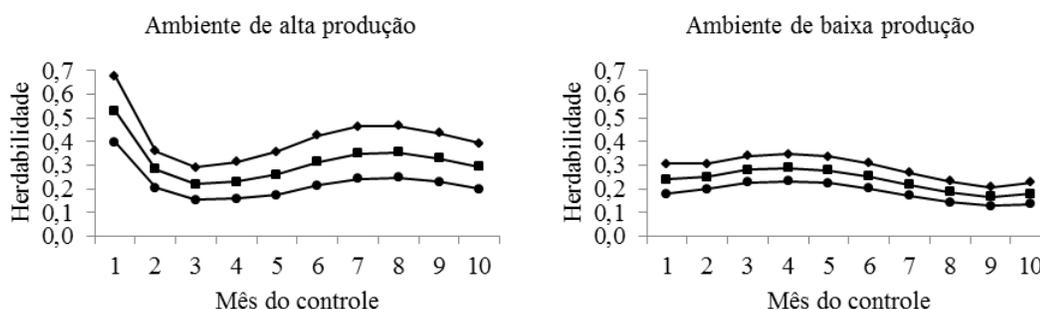


Figura 2. Estimativa de média e intervalo de alta densidade a 95% das distribuições posteriores da herdabilidade para os diferentes meses de controle, para os dois ambientes de produção.

Independentemente do grupo formado, o comportamento das correlações genéticas entre as produções de leite foram menores conforme se distanciavam os controles, sendo de maior magnitude nos períodos homólogos e adjacentes, em ambos os grupos. Todas as estimativas para as correlações genéticas nos dois ambientes foram positivas e superiores a 0,22, indicando que a seleção para esta característica, em determinado ponto da lactação de acordo com o ambiente, terá reflexo positivo sobre todos os outros pontos da lactação.

As correlações de ambiente permanente apresentaram valores altos e positivos entre controles próximos, mas à medida que aumentou a distância entre controles estas diminuíram até o ponto em que se tornaram negativas, variando de -0,24 até 0,97. Bignardi et al. (2009), em bovinos da raça Holandesa, reportaram que as correlações de ambiente permanente tendem a ser negativas no final da lactação, além de sugerirem que os polinômios de Legendre apresentam uma maior ênfase nas observações dos extremos da curva de lactação, o qual gera um ajuste menos adequado causando dificuldade ao modelar PLDC.

Por outro lado, as correlações genéticas para ambiente de alta produção variaram de 0,22 a 0,99 e no ambiente de baixa produção seu comportamento variou de 0,54 a 0,99. No caso das correlações entre ambientes variaram de 0,29 a 0,92, indicando a presença do efeito de interação genótipo-ambiente.

Conclusões

A técnica de agrupamento não hierárquico de *k-means* evidenciou a diferenciação entre os grupos ambientais descritos, com valores maiores para os parâmetros no ambiente de alta produção. Houve indicativo de interação genótipo-ambiente como efeito direto de heterogeneidade de variâncias entre ambientes.

Literatura citada

- BIGNARDI, A.; EL FARO, L.; CARDOSO, V.; MACHADO, P.; ALBUQUERQUE, L. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in Southeastern Brazil. *Livestock Science*, v. 123, p. 1–7, 2009.
- HERRERA, L.G.; EL FARO, L.; GALVÃO DE ALBUQUERQUE, et al. Estimativas de parâmetros genéticos para produção de leite e persistência da lactação em vacas Gir, aplicando modelos de regressão aleatória. *Revista Brasileira de Zootecnia*, v.37, p.1584-1594, 2008.
- MISZTAL, I. 2010. THRGIBBSF90 family of programs. Available at: <<http://nce.ads.uga.edu/~ignancy/newprograms.html>>.
- SAVEGNAGO, R.P.; EL FARO, L.; CAETANO, S.L. et al. Análise de agrupamento hierárquico e não hierárquico dos valores fenotípicos e genéticos da produção de leite de vacas Holandesas em primeira e segunda lactação. In: 57ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBRAS). Piracicaba, SP. Anais. 2012.
- VICINI, L.; SOUZA, A.M. Análise multivariada da teoria à prática. *UFMS, CCNE, Santa Maria*, 2005.