

Modelos de regresiones aleatorias para la estimación de parámetros genéticos y estudios de curvas de lactancia del Holstein en Cuba

Lucía Fernández¹, H. Tonhati², L.G. Albuquerque², R. R. Aspilcueta-Borquis² y A. Menéndez Buxadera³

¹ Universidad Agraria de La Habana "Fructuoso Rodríguez Pérez"

² Universidad Estadual Paulista, Campus de Jaboticabal, Brasil

³ Centro de Control Pecuario, Ministerio de la Agricultura, Cuba
Correo electrónico: lucia@isch.edu.cu

Para estimar parámetros genéticos, funciones de covarianza y curvas de lactancia en producciones de leche en el día de control, se analizaron 2861 pesajes individuales de la primera lactancia de 357 vacas de la raza Holstein, pertenecientes a una empresa de la provincia La Habana, Cuba, durante el período 1999-2008. Se compararon diez modelos diferentes de Regresiones Aleatorias (MRA) con la aplicación de polinomios ortogonales de Legendre. Se incluyó entre los efectos fijos el efecto combinado rebaño-año de pesaje-mes de pesaje y la curva fija de orden 4. Como aleatorios se consideró el efecto genético aditivo y el de ambiente permanente. La edad al parto se valoró como covariable. Se concibieron 31 clases de días en lactación (con intervalo de 10 d), para un total de 310 d. El modelo de Ali y Schaeffer fue el que mejor caracterizó la curva de lactancia con efectos incluidos (con ligero pico a los 35 d, con 10.66kg/d). La mejor estructura de varianza fue la de orden 5. El modelo de regresión aleatoria más adecuado fue el que consideró polinomio de Legendre de grado 3 para el efecto genético aditivo, y grado 6 para el de ambiente permanente. Los mayores valores de heredabilidad (0.12-0.26) se alcanzaron entre los 20 y 145 d. Las correlaciones genéticas oscilaron entre 0.52 -1, y fueron más fuertes al inicio y centro de la lactación. Se concluye que los MRA constituyen una vía eficiente para estimar parámetros y variaciones genéticas. En este estudio, el uso de varianzas residuales heterogéneas resultó más adecuado para modelar la producción de leche en el día de control.

Palabras clave: *regresiones aleatorias, día de control, polinomios de Legendre, Holstein.*

Los programas de mejoramiento genético se han incrementado notablemente en la actualidad. Su éxito depende, entre otras cosas, de la exactitud con que se realice la evaluación genética de los animales. Esta depende de la estructura, distribución de los datos, modelo y método que se utilice en la evaluación. Autores como Ptak y Schaeffer (1993) y Swalve (1995) han sugerido los modelos de regresión aleatoria (MRA) para evaluaciones genéticas de datos longitudinales, como la producción de leche en el día de control. Estos modelos consideran, entre otros aspectos, la forma de la curva de lactancia, el efecto particular del día de control para todas las vacas y los efectos específicos para cada vaca en el día de control.

Los trabajos de El Faro y Albuquerque (2005), González *et al.* (2008), Costa *et al.* (2008) y Sesana (2008), entre otros, corroboran la eficiencia de estos modelos en bovinos y bubalinos lecheros. En Cuba, a pesar de los esfuerzos realizados, existen limitaciones en cuanto a la utilización de los modelos de regresión aleatoria para producciones de leche en el día de control, debido a las complejidades propias de estos métodos, como el uso de las genealogías, la recopilación de la información y su procesamiento. El objetivo de este trabajo fue estimar los parámetros genéticos, funciones de covarianza y curva de lactancia, a partir de producciones de leche en el día de control, con la aplicación de modelos de regresión aleatoria mediante la utilización de polinomios ortogonales de Legendre.

Materiales y Métodos

Se analizó un total de 2 861 pesajes individuales de la primera lactancia, de 357 vacas Holstein, pertenecientes a una empresa ganadera, ubicada al oeste de la provincia La Habana, Cuba. Los datos correspondieron al período 1999-2008, distribuidos en 24 rebaños. Representaron la progenie de 60 sementales, hijas de 294 reproductoras, donde el fichero pedigrí incluyó el animal que produce el registro, madre y padre, para un total de 1329 animales. Las producciones de leche en el día de control se obtuvieron a partir de los cinco días. Se contó, al menos, con cinco pesajes por animal. Se construyeron 493 grupos contemporáneos, según rebaño- año de pesaje-mes de pesaje, con más de cinco observaciones por grupo, y producciones dentro de la media + 3 desviación estándar.

Se modeló la curva de lactancia con efectos incluidos en los 31 intervalos. Se probaron tres modelos matemáticos:

- 1- Función gamma incompleta (Wood)

$$Y_t = a t^b e^{-c t}$$
- 2- Regresión múltiple (Ali y Schaeffer)

$$Y_t = a + b \frac{t}{305} + c \left(\frac{t}{305} \right)^2 + d \cdot h \frac{305}{t} + e \left(h \frac{305}{t} \right)^2$$
- 3- Cuadrática logarítmica

$$y_t = a + b t + c t^2 + d h t$$
 donde :
 Yt - producción promedio de leche en el día t
 a , b , c , d , e - parámetros a estimar

$e \approx 2.7182$ base de los logaritmos naturales

ln - logaritmo natural

t -días de lactancia

En el procesamiento de la información se utilizó el software Statgraphics Plus, versión 5.1. La estimación de los parámetros de modelos no lineales requirió de métodos iterativos. En la validación de los modelos se incluyó el coeficiente de determinación (R^2) y el coeficiente de determinación ajustado (R^2A). Este último por ser modelos comparados, con diferentes números de parámetros. Se incluyó además, el error estándar de estimación, así como la media absoluta del error y la suma de cuadrado del error. Se aplicó la dócima de Durbin Watson (DW) para verificar el cumplimiento de los supuestos teóricos.

Se compararon diez modelos diferentes de regresión aleatoria, que utilizaron el programa Wombat para la estimación de los valores genéticos (Meyer 2008) mediante métodos de máxima verosimilitud restringida, y los polinomios ortogonales de Legendre. Se incluyó entre los efectos fijos, el combinado rebaño-año de pesaje-mes de pesaje, y la curva fija de orden 4 para modelar la tendencia media de la población. Como efectos aleatorios, se incluyó el genético aditivo, el de ambiente permanente de animal y el error (se compararon varianzas homogéneas y heterogéneas, con diferentes números de clases). Se incorporó además como covariable, la edad al primer parto (a partir de la expresión cuadrática). Se consideraron solo edades al parto, entre 24 y 48 meses. A partir de los días en lactación, se construyeron 31 clases (con intervalo de 10 d), para un total de 310 d.

Para la conformación de estos modelos se tuvieron en cuenta las varianzas homogéneas y heterogéneas, de 5, 7, 12 y 31 órdenes, así como la combinación de polinomios de diferentes órdenes para el efecto genético aditivo y de ambiente permanente. El fichero pedigri incluyó al animal que produjo el registro, la madre y el padre.

Descripción general del modelo de regresión aleatoria:

$$Y_{ij} = F + \sum_{m=0}^{K_b-1} \beta_m \phi_m(t_i) + \sum_{m=0}^{k_a-1} \alpha_{jm} \phi_m(t_{ij}) + \sum_{m=0}^{K_{ab}-1} \gamma_{jm} \phi_m(t_{ij}) + e_{ij}$$

Donde:

Y_{ij} - producción de leche en el día de control

F - conjunto de efectos fijos (grupo contemporáneo)

β_m - conjunto de m regresores fijos para la curva media de población

$\phi_m(t_i)$ - función de regresión de orden k_b , que describe la curva media de población, de acuerdo con el tiempo de lactación (cada 10 d) (modelada por polinomios ortogonales de Legendre de tercer orden)(li)

α_{jm} - regresores aleatorios genético aditivo para cada animal

$\phi_m(t_{ij})$ - funciones de regresión, que describen la trayectoria de cada individuo j, de acuerdo con el tiempo de lactación (cada 10 d)(ti) para los efectos aleatorios genético aditivo y de ambiente permanente

Revista Cubana de Ciencia Agrícola, Tomo 45, Número 1, 2011.

γ_{im} - regresores aleatorios de ambiente permanente para cada animal

k_b k_a k_{ab} - órdenes del polinomio para describir la curva media y el efecto genético aditivo y de ambiente permanente, respectivamente

e_{ij} - error aleatorio asociado a cada control y del animal (i)

Para la bondad de ajuste y selección del mejor modelo, se utilizaron los criterios de Akaike (AIC) y Bayesiano de Schears (BIC), según Wolfinger (1993), a partir del procesamiento Wombat (Meyer 2008).

Se analizó el comportamiento de varianzas fenotípicas, aditiva, de ambiente permanente y residual, estimadas a partir del mejor modelo de regresión aleatoria, así como la heredabilidad por clase. Se estimaron las heredabilidades, mediante la aplicación test-day model, ambos resultados se compararon. Se estimaron y analizaron correlaciones genéticas en el mejor modelo.

Resultados y Discusión

Los niveles promedio de producción de leche en el día de control fueron de 8.37 ± 0.06 kg, con coeficiente de variación 42.38 %. Se considera que uno de los aspectos que influyó en los bajos niveles de producción y en la alta variabilidad fue el trabajo en condiciones de explotación comercial y con animales que producen, fundamentalmente, en condiciones de pastoreo (con limitados suplementos adicionales), pero estos promedios han sido superados en Cuba con mejores niveles de alimentación (tipos de pastoreo). Esto se puede corroborar en los trabajos de Michelena *et al.* (2009), quienes informaron promedios diarios de leche entre 9.97 kg y 14.5 kg en Holstein.

Las medias de las producciones de leche (kg/d) para cada intervalo se presentan en la figura 1, con valores promedio que oscilaron entre 5.11 ± 0.47 kg y 11.01 ± 0.34 kg, con ligero incremento en la producción, aproximadamente a los 30 d de lactancia.

Se estimó la curva de lactancia con efectos incluidos en los 31 intervalos. De los tres modelos probados, el de Ali y Schaerffer fue el más apropiado (tabla 1), con mayor coeficiente de determinación ($R^2 = 99.91$) y menor error estándar de estimación, media absoluta del error y suma de cuadrado residual.

La figura 2 muestra la curva de lactancia, donde el pico se alcanzó, aproximadamente, a los 35 d, con una producción de 10.66 kg/d. Posteriormente se manifestó una tasa de declinación.

Los resultados que se muestran en la tabla 2 se obtuvieron de la comparación de los modelos de regresión aleatoria, con diferentes órdenes de polinomios para efectos genéticos aditivos y de ambiente permanente, así como varianzas homogéneas y heterogéneas de distintos órdenes, y la curva fija de orden 4 para modelar la tendencia media de la población.

La tabla 2 presenta también los valores de AIC.

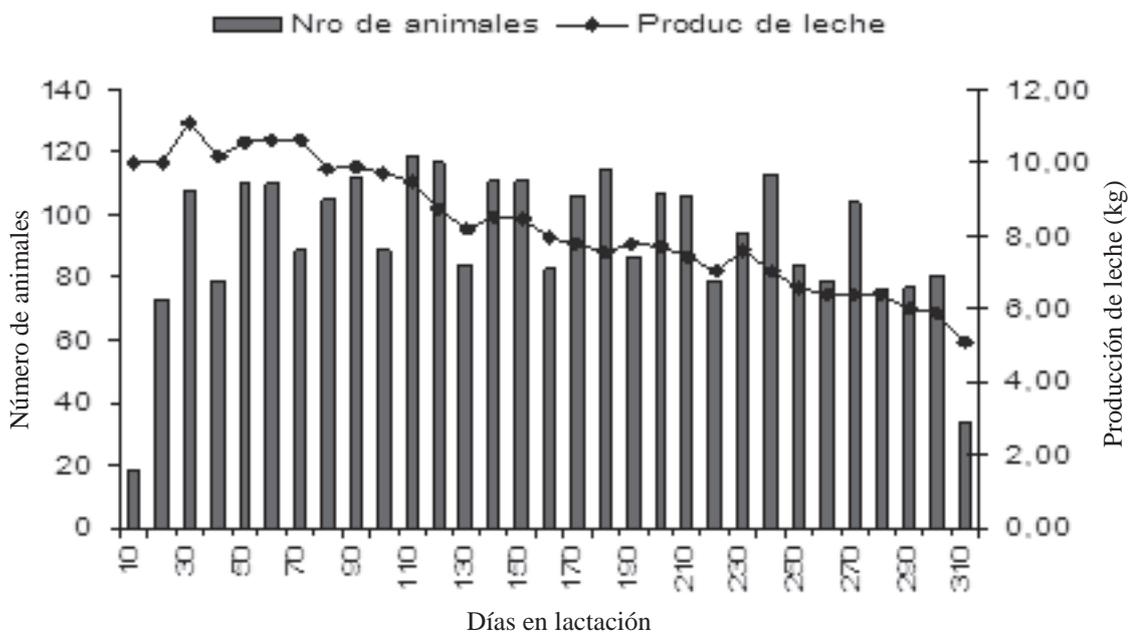
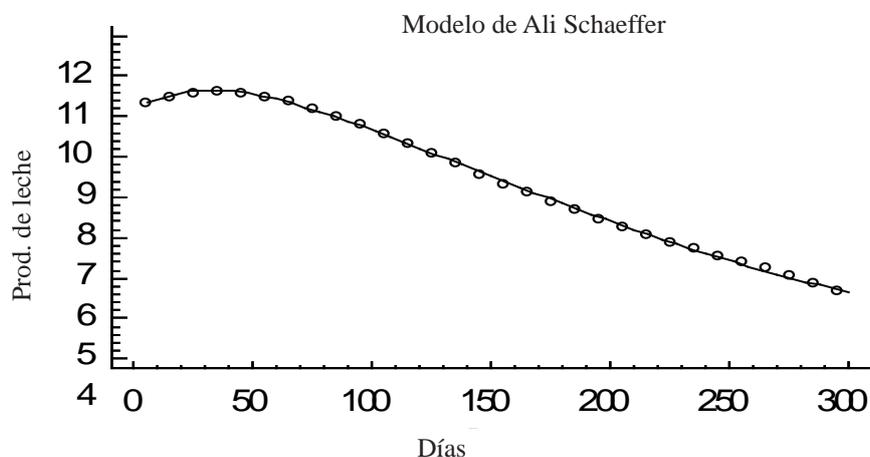


Figura 1. Distribución de frecuencia de las producciones promedio de leche (kg/d), con intervalo de 10 d y número de observaciones por clase.

Tabla 1. Resultados de los ajustes de los modelos de regresión a la curva de lactancia

Criterios estadísticos	Modelos		
	Wood	Ali y Schaeffer	Cuadrática logarítmica
Método de ajuste de los modelos	Marquardt	MCO	MCO
Significación del modelo	0.0001	0.0001	0.0001
Coefficiente de determinación (R ²) %	92.27	99.91	99.74
Coefficiente de determinación ajustado (R ² A)%	91.72	99.89	99.71
Error estándar de estimación	0.50	0.05	0.09
Media absoluta del error	0.31	0.04	0.07
Suma de cuadrado del error	7.06	0.07	0.23
Incorrelación (Dócima de Durbin Watson)	0.38	0.34	0.73

MCO Método de los Mínimos Cuadrados Ordinarios



$$Y_t = 19.36 - 19.53 \cdot \frac{t}{305} + 5.75 \cdot \left(\frac{t}{305}\right)^2 - 4.02 \cdot \ln \frac{305}{t} + 0.46 \left(\ln \frac{305}{t}\right)^2$$

Figura 2. Representación de la curva de lactancia a partir del mejor modelo.

Tabla 2. Resumen de los diez modelos de regresiones aleatorias y criterios de selección

Modelo	Ka	Kp	c	p	2 LOG L	AIC	BIC
1	3	3	hom	13	-3340.42	6706.84	6781.84
2	3	3	het 31	48	-3305.03	6696.06	6944.15
3	3	3	het 12	24	-3311.92	6671.85	6810.31
4	3	3	het 7	19	-3322.94	6683.87	6793.49
5	3	3	het 5	17	-3324.91	6683.83	6781.91
6	3	4	het 7	21	-3316.59	6675.17	6796.33
7	3	5	het 7	26	-3308.99	6669.98	6819.99
8	3	6	het 7	32	-2897.24	5858.47	6043.08
9	4	4	het 7	25	-3316.35	6682.69	6826.93
10	4	5	het 7	30	-3308.24	6676.48	6849.54

c-classes de varianzas (het) y homogeneidad de varianzas (hom)

Al agrupar los modelos del 1 al 5, el que tuvo peor ajuste fue el de varianza homogénea. El más apropiado fue el de varianza heterogénea de 5 órdenes (modelo 5).

Una vez seleccionada la estructura de varianzas residuales, se comparó el resto de los modelos (6-10), cuyos valores de AIC oscilaron entre 5858.47 y 6706.84. En el caso del criterio BIC, que es más riguroso en cuanto a la consideración del número de parámetros, los valores estuvieron entre 6043.08 y 6781.84. Para ambos criterios, el modelo 8 resultó ser el más adecuado para describir la variación de la producción de leche en las condiciones de este estudio.

Las varianzas fenotípicas, genética aditiva, de ambiente permanente y residuales para el modelo 8 aparecen en la figura 3. Las varianzas fenotípicas fueron más altas en los primeros 20 d, y posteriormente se estabilizaron hasta el final de la lactación. Esta irregularidad que se encontró al

inicio de la lactancia coincide con lo informado por El Faro *et al.* (2005), Costa *et al.* (2008) y González *et al.* (2008). Se considera que este comportamiento puede estar sujeto al número de observaciones en el primer control. La varianza residual presentó la misma tendencia que la fenotípica, pero en menor magnitud. La varianza aditiva mostró decrecimiento en toda su trayectoria, y la aditiva permanente fue constante en toda la lactación. Estos resultados, de forma general, siguen la misma tendencia que refiere la literatura especializada en bovino de leche y bubalinos.

Al comparar la varianza genética con lo obtenido en el análisis de dimensión finita (figura 3), el modelo de regresión aleatoria (MRA) estimó mayores valores de varianza genética aditiva, con respecto a lo obtenido en el pesaje en el día de control, además de presentar un comportamiento más estable. En la varianza residual y

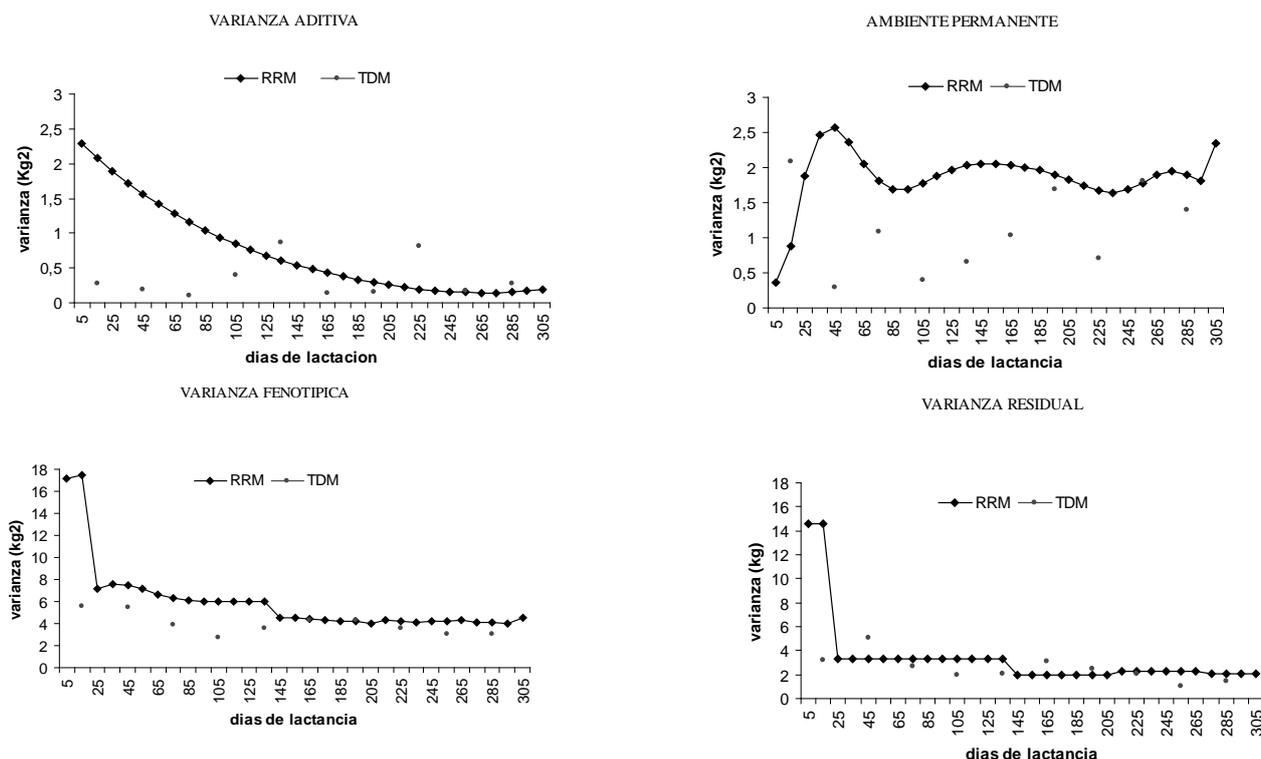


Figura 3. Varianzas fenotípica, aditiva, de ambiente permanente y residual, estimadas a partir del modelo 8 (MRA), y varianzas obtenidas en el análisis de dimensión finita (TDM)

fenotípica se presentaron similares resultados por ambos procedimientos, mientras que la de ambiente permanente se mostró más estable y con mejor comportamiento en el modelo de regresión aleatoria. Estos elementos permiten encontrar mayores valores de heredabilidad en los modelos de regresión aleatoria.

Según los resultados que se muestran en la figura 4, los mayores valores de heredabilidad (0.12-0.26) se alcanzaron entre los 25 y 145 d. Los resultados de El Faro y Albuquerque (2005) y Sesana (2008), con modelos de regresiones aleatorias similares al de este trabajo, mostraron patrones semejantes, aunque con valores de heredabilidad superiores, al inicio y final de la lactación. En este estudio no se manifestó ese crecimiento en los valores de h^2 al final de la lactancia, debido posiblemente al tamaño de la muestra utilizada, lo que puede tener un efecto determinante en las varianzas estimadas. En este sentido, se ha planteado que los componentes de varianza dependen de la estructura y de la cantidad de datos efectivos, por lo que se recomienda emplear datos de no menos de

10 000 observaciones para estimaciones adecuadas (van der Werf y Goddard 2003).

Los valores de heredabilidad, estimados a partir del test-day para clases mensuales, se representan en la figura 4. Estos fueron irregulares, y los estimados por MRA fueron más elevados en el quinto y octavo mes. Sin embargo, en el resto de la lactancia estos valores de h^2 resultaron, generalmente, mucho menores que los estimados para MRA.

El estudio permitió identificar que es posible realizar la selección de animales Holstein en Cuba, preferiblemente próximos al pico de producción, pues es el momento en que alcanzan los mayores valores de heredabilidad.

Las estimaciones de las correlaciones genéticas para las producciones de leche en diferentes clases (días de lactancia/periodo=10) se muestran en la figura 5. En todos los casos, se manifiesta el patrón clásico, donde las correlaciones son más débiles a medida que aumentan los días en lactación. Se debe destacar que en la mayoría de las clases se presentaron valores de correlación por encima de 0.90.

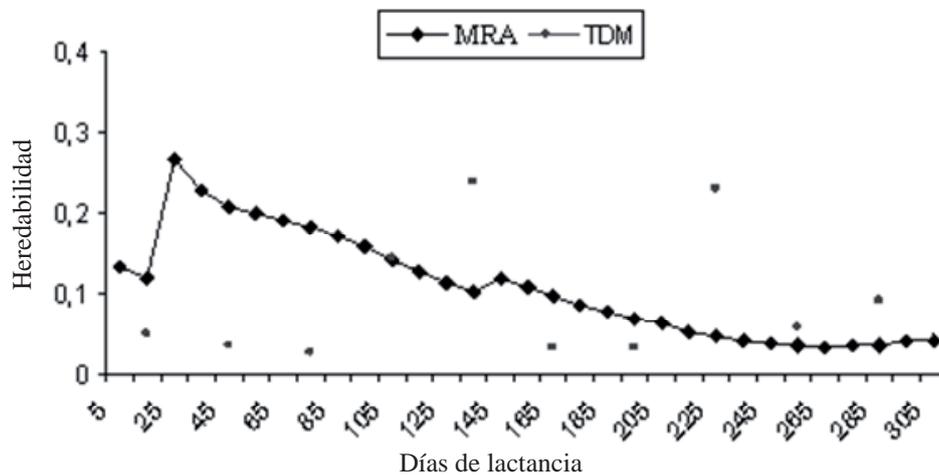


Figura 4. Estimación de heredabilidades para las producciones de leche en el día de control para MRA y test-day model (REMLF90).

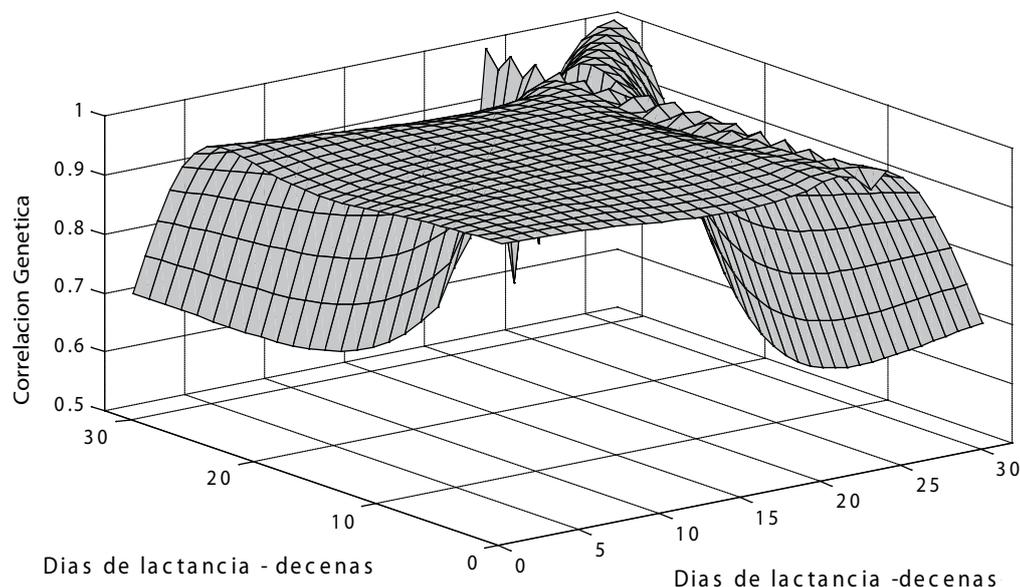


Figura 5. Correlaciones genéticas entre las producciones de leche en el día de control por clases.

Se concluye que los modelos de regresiones aleatorias, con el empleo de polinomios de Legendre, constituyen una vía eficiente para estimar parámetros y variaciones genéticas. El uso de varianzas residuales heterogeneas resultó más adecuado para modelar la producción de leche en el día de control. La mejor estructura de varianza fue la de orden 5.

En las condiciones de este estudio, el modelo de regresión aleatoria más adecuado para describir la variación de la producción de leche fue el que consideró polinomio de Legendre de grado 3 para el efecto genético aditivo, y de grado 6 para el efecto de ambiente permanente. El modelo que mejor caracterizó el comportamiento de la curva de lactancia fue el de Ali y Schaeffer.

Agradecimientos

Se agradece a la Empresa Pecuaria Genética “Los Naranjos”, en la provincia La Habana, por la información brindada (bases de datos) para la realización del presente trabajo, así como a la Coordinación de Perfeccionamiento de Personal de Nivel Superior (CAPES) y a la Universidad Estadual Paulista (UNESP) por el apoyo, formación e intercambio en los análisis realizados.

Referencias

Costa, C.N., Rodríguez, C.M., Packer, I.U., Ferreira, A., Teixeira, N.M. & Araújo, J. 2008. Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. *R. Bras. Zootec.*v.37 n. 4.

El Faro, L. & Albuquerque, L.G. 2005. Predição de valores genéticos para a produção de leite no dia de controle e para

Revista Cubana de Ciencia Agrícola, Tomo 45, Número 1, 2011. a produção acumulada ate 305 dias. *R.Bras.Zootec.*v.34 n. 2.

Gonzalez, L.G., El Faro, L., Albuquerque, L.G., Tonhati, H. & Cavallari, C.H. 2008. Estimativas de parâmetros genéticos para produção de leite e persistência de lactações em vacas gir, aplicando modelos de regressão aleatória. *R. Bras. Zootec.*37:1584

Meyer, K. 2008. Wombat. A program for Mixed Model Analices by restricted Maximun Likelihood. University of New England. Australia

Michelena, J. B., Ruiz, T., Martínez, R. & Padilla, C. 2009. Producción de leche a bajo costo Disponible: <http://www.fira.gob.mx:8081/sas/docs/MemoriasEventos/ForoTuxpan/Producci%C3%B3n%20de%20Carne%20Vacuna%20a%20Bajo%20Costo.pps>. Consultado: marzo de 2009.

Ptak, E. & Schaeffer, L.R. 1993. Use of test day yield for genetic evaluation of dairy sires and cows. *Livestock Production Science.* 34:23

Sesana, R.C. 2008. Estimativas de parâmetros genéticos para a produção de leite em búfalas por modelos de repetibilidade, multi-característica e de regressão aleatória. Tesis apresentada em opção a Mestre em Genética e Melhoramento Animal. Jaboticabal. São Paulo, Brasil.

Swalve, H. 1995. Test day models in the analysis of dairy production data-a review. *Arch. Tierz. Dummerstorf.* 38:591

Van der Werf, J. & Goddard, M. 2003. Models and Methods for Genetic Analysis. Course Notes. University of New England (UNE).

Wolfinger, R. 1993. Covariance structure selection in general mixed models, *Communications in Statistics.* 22:1079

Recibido: 11 de septiembre de 2009