

IMPLEMENTACIÓN DE MODELOS DE REGRESIÓN ALEATORIA PARA CARACTERES DE CRECIMIENTO EN ANIMALES CRUZADOS EN EL SUR DEL CESAR, COLOMBIA

IMPLEMENTATION OF RANDOM REGRESSION MODELS FOR GROWTH TRAITS IN CROSSBRED CATTLE AT SOUTHERN CESAR COLOMBIA

Regresión aleatoria en crecimiento de animales cruzados

Martínez C.A.^{1*}, Manrique C.¹, Jiménez A.^{1,2}, Coy B. Y.¹, Elzo M. A.³

¹Grupo de estudio en mejoramiento y modelación animal GEMA, Departamento de producción animal, Universidad Nacional de Colombia, Bogotá D.C. Colombia. *camartinezn@unal.edu.co

²Asociación Colombiana de criadores de Ganado Cebú ASOCEBU, Bogotá D.C.

³University of Florida, Department of Animal Sciences, Gainesville, FL

Palabras clave:

Funciones de covarianza
Genética
cuantitativa

Keywords:

Covariance
functions
Quantitative
genetics

Abstract

Under the classic theory of quantitative genetics, the repeated measurements taken in the same individual at different times have been studied using repeatability models in which it is assumed that the repeated measurements correspond to the same character, or using multiple trait models in which each measurement is considered a different trait. One alternative is the use of random regression models and covariance functions in which the traits are modeled as a continuous time function through orthogonal functions. The objective of this work was to implement random regression models and covariance functions using Legendre polynomials in quantitative genetics for growth longitudinal data in cattle. Homogeneous and heterogeneous variances random regression models and a repeatability model were compared. The results of this research showed a better adjust of the heterogeneous variance random regression models.

Resumen

Bajo la teoría clásica de genética cuantitativa las mediciones repetidas tomadas de un mismo individuo en distintos tiempos se han estudiado bien sea con modelos de repetibilidad, en los cuales se asume que las medidas repetidas corresponden al mismo carácter, o mediante modelos de múltiples caracteres, en los cuales cada medición se considera una característica diferente. Una alternativa son los modelos de regresión aleatoria y funciones de covarianza en los cuales los registros se modelan como una función continua del tiempo mediante funciones ortogonales. El objetivo del presente trabajo fue implementar modelos de regresión aleatoria y funciones de covarianza empleando polinomios de Legendre en genética cuantitativa para datos longitudinales de crecimiento en bovinos. Se compararon modelos de regresión aleatoria con varianzas residuales homogéneas y heterogéneas y un modelo de repetibilidad. Los resultados de esta investigación mostraron un mejor ajuste de los modelos de regresión aleatoria con varianzas residuales heterogéneas.

Introducción

En la evaluación genética tradicional se obtienen los mejores predictores lineales insesgados (MPLI) del valor genético de los individuos en una población para uno o varios caracteres mediante el uso de modelos mixtos (Henderson, 1974). Para obtener dichos valores se requiere la estimación de componentes de varianza. Uno de los procedimientos más empleados para estimarlos es el de máxima verosimilitud restringida (MVR) (Patterson & Thompson, 1971). Cuando se tienen datos tomados a través del tiempo para varios individuos, se tiene una estructura de datos longitudinales, la cual merece un tratamiento estadístico especial (Verbeque & Molenberghs 2000). Varias formas de tratar estos datos han sido propuestas. Los modelos para caracteres repetidos pueden clasificarse en dos grupos, de dimensión finita y de dimensión infinita (Meyer & Hill 1997). El primer tipo de modelos evalúa rasgos a una edad determinada, en estos se asume que la variación es discreta. Si bien los resultados de estos modelos pueden ser satisfactorios la variación de caracteres como el peso es continua. Así, en los modelos de dimensión infinita se modela el fenotipo como función continua del tiempo, (Kirpatrick et. al. 1990). Las Funciones de covarianza (FC) y los modelos de regresión aleatoria (MRA) inicialmente fueron

descritos por aparte. Pero trabajos como el de Meyer & Hill (1997) mostraron la equivalencia de las FC y un caso especial de los MRA, al ver que bajo el supuesto de residuales independientes la estructura de covarianza de los coeficientes de regresión aleatorios podía ser modelada mediante una FC como la descrita por Kirpatrick et. al. (1990). Desde entonces se les llamó modelos de regresión aleatoria y funciones de covarianza (MRA-FC). El objetivo de este trabajo fue implementar MRA-FC para estimar componentes de varianza en una población bovina multirracial en Colombia. Se compararon 2 modelos de regresión aleatoria y un modelo de repetibilidad simple mediante tres criterios de información: Criterio de información de Akaike (AIC), criterio de información de Akaike corregido (AICC) y el criterio de información Bayesiano de Schwarz (BIC).

Material y métodos

Se trabajó con 1090 registros de peso vivo a diferentes edades facilitados por la Asociación Colombiana de Criadores de Ganado Cebú ASOCEBÚ. La población estuvo conformada por 352 progenies de 37 toros de 9 razas y vacas Brahman gris ubicadas en el sur del Cesar, Colombia. Se emplearon 3 modelos. El modelo 1 correspondió al modelo de repetibilidad, el modelo 2 fue un MRA-FC empleando polinomios de Legendre de grado 2 para modelar la varianza genética aditiva ($k_a = 3$), grado 3 para modelar la varianza de ambiente permanente ($k_p = 4$) y varianza residual homogénea, el modelo 3 fue un MRA-FC con grado de polinomio igual al modelo 2 pero con varianzas residuales heterogeneas a través de 4 clases de intervalos: 1 a 120 d, 121 a 240 d, 241 a 365 d y 366 a 492 d. Estos intervalos se definieron teniendo en cuenta etapas de la vida del animal en las cuales se dan eventos de importancia biológica. Así, la varianza del residual es una función escalonada con 4 sub-dominios. Los efectos fijos fueron: grupo contemporáneo (subclase año-época-sexo-finca), edad del animal, fracción esperada de cada raza en el animal, y heterosis (probabilidad de alelos diferentes en un locus). Los efectos aleatorios fueron: efectos genético aditivo del animal y de ambiente permanente. Los modelos se compararon mediante el criterio de información bayesiana de Schwarz (BIC), el criterio de información de Akaike (AIC) y el criterio de información de Akaike corregido el cual tiene en cuenta el tamaño de la muestra (AICC) (Littel et. al. 2006). Una vez seleccionado el modelo con menores criterios de información, se estimaron los valores y vectores propios de la matriz de covarianzas estimada, con los vectores propios se calcularon las funciones propias, que son el equivalente en los modelos de dimensión infinita a los vectores propios. Las funciones propias, son funciones continuas que representan una posible deformación evolutiva de la trayectoria media de crecimiento (Kirpatrick et. al. 1990). Estas se calcularon así: $\psi_i(a) = \sum_{j=0}^2 \{c_{\psi_i}\}_j \phi_j a^*$, donde $\{c_{\psi_i}\}_j$ corresponde a la j-ésima entrada del i-ésimo vector propio de la matriz de coeficientes de la FC. Las estimaciones MVR de componentes de covarianza se realizaron empleando el programa Wombat (Meyer 2007) empleando el algoritmo Average Information (AI).

El MRA-FC empleado fue:

$$y = X\beta + \Phi_a a + \Phi_p p + e$$

$$\text{Var} \begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} K_a \otimes A & 0 & 0 \\ 0 & K_p \otimes I & 0 \\ 0 & 0 & R \end{bmatrix}$$

$$\text{Var}(y) = \Phi_a K_a \otimes A \Phi_a' + \Phi_p K_p \otimes I \Phi_p' + R$$

Donde y es el vector aleatorio que contiene los valores realizados de la variable aleatoria peso ordenados por animal, β es un vector desconocido de efectos fijos, a y p son vectores aleatorios desconocidos que contienen los coeficientes de regresión aleatoria de los efectos genéticos aditivos y de entorno permanente respectivamente, e es el vector de residuales, X es la matriz de incidencia que relaciona los elementos de β al vector de registros mediante unos, Φ_a y Φ_p son las matrices de incidencia conocidas de los vectores a y p las cuales contienen los polinomios ortogonales evaluados en las edades a las cuales se tomaron los registros; K_a y K_p son las matrices de coeficientes de las funciones de covarianza genética aditiva y de entorno permanente, A es la matriz de parentesco, I es la matriz de identidad y \otimes representa el producto directo o de Kronecker. En este caso el logaritmo de la función de verosimilitud (L) fue:

$$\ln(L) = -1/2 [\text{const} + n \ln |\Phi_p K_p \Phi_p' + R| + n_a \ln |K_a| + n_a \ln |\Phi_a \Phi_a'| + t \ln |A| + \ln |C| + y'Py]$$

Const. Son las constantes de la función, t el número de edades a las cuales se tomaron mediciones, C es la matriz de coeficientes de las ecuaciones de modelos mixtos, y'Py es la suma de cuadrados de los residuales.

En general, las ecuaciones de modelos mixtos fueron:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 \\ Z_1'R^{-1}X & Z_1'R^{-1}Z_1 + A^{-1} \otimes K_a^{-1} & Z_1'R^{-1}Z_2 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z_2'R^{-1}Z_2 + I \otimes K_p^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ a \\ p \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \end{bmatrix}$$

Cuando el orden del polinomio es cero, es decir solo se tiene el término independiente, el MRA-FC es equivalente al modelo de repetibilidad simple (Meyer & Hill 1997). Por tanto estas ecuaciones también aplica a este modelo dado que es un caso especial de los MRA-FC.

Resultados y discusión

De acuerdo a la tabla 1, los resultados indicaron que el mejor modelo fue el 3 presentó un mejor ajuste, pues sus valores de BIC, AIC y AICC fueron los menores (Littel et. al. 2006).

Tabla I. Rango de edades y estadísticas descriptivas de cada momento de medición (*Age range and descriptive statistics for each measurement time*)

Modelo	AIC	AICC	BIC
1	2677.70	2677.71	2681.43
2	1814.17	1814.32	1835.34
3	1787.97	1788.17	1812.57

En notación matricial la función de covarianza genética aditiva (FCGA) fue:

$$CGA(a_1^*, a_2^*) = [\phi_0(a_1^*) \quad \phi_1(a_1^*) \quad \phi_2(a_1^*)] \begin{bmatrix} 489.5 & 272.2 & 35.9 \\ 272.2 & 159.7 & 24.3 \\ 35.9 & 24.3 & 4.8 \end{bmatrix} \begin{bmatrix} \phi_0(a_2^*) \\ \phi_1(a_2^*) \\ \phi_2(a_2^*) \end{bmatrix}$$

De acuerdo a esta función es que la varianza y las covarianzas genéticas aditivas aumentan a medida que los animales crecen. Por lo tanto, las covarianzas de los pesos a mayores edades son más grandes que aquellas entre pesos en la fase inicial del crecimiento. El primer valor propio de la matriz de coeficientes de la función de covarianza genética aditiva fue 643.9 y este explicó el 98.76% de la varianza genética aditiva total. El vector propio asociado a este valor propio fue: [0.87 0.49 0.07]'. La primera función característica o función propia fue: $\widehat{\psi}_1(a) = 0.56 + 0.6a^* + 0.39a^{*2}$. La figura 1 muestra la gráfica de esta función. Lo que muestra esta función es que la selección a cualquier edad aumentará los pesos en las demás edades, especialmente en las edades mayores. Y dado el alto porcentaje de la varianza genética aditiva que explica el valor propio asociado a esta función, dicha selección producirá una notable modificación en la trayectoria media de crecimiento de la población. Por lo tanto, se deben seleccionar individuos más pesados a edades tempranas, lo cual aumentará los pesos de los animales a edades mayores, como por ejemplo a los 365 días, que es una edad de importancia comercial.

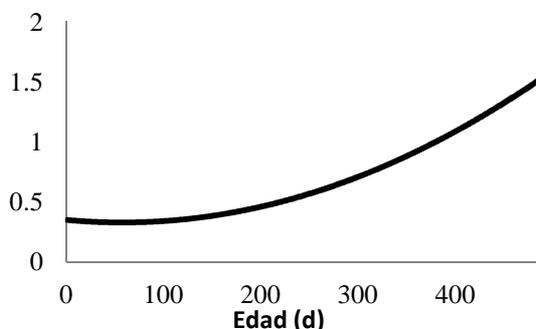


Figura 1. Primera función propia de la FCGA (*First eigenfunction of FCGA*)

Conclusiones

Los MRA-FC son una herramienta valiosa en los análisis genéticos de datos longitudinales, ya que desde el punto de vista tanto estadístico como biológico describen de manera adecuada datos longitudinales como los de

crecimiento. Los resultados de este trabajo sugieren que los modelos de repetibilidad no son adecuados para describir la trayectoria del crecimiento y que los MRA-FC presentan una mejor descripción de los datos cuando se modelan varianzas residuales heterogéneas a través de intervalos de la trayectoria.

Bibliografía

- Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario, Canada.
- Patterson HD, and Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979-993.
- Littell R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. SAS® for mixed models, 2nd Ed. SAS Institute Inc., Cary.
- Meyer, K. and W.G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or “repeated” records by restricted maximum likelihood. *Livest Prod Sci* 47, 185-200.
- Meyer, K. 2007. WOMBAT –A program for mixed models analyses in quantitative genetics by REML. *J Zhejiang Univ Sci* 8, 815-821.
- Verveque, G., and G. Molenrghs. 2000. Linear mixed models for longitudinal data, 1st Ed, Springer, New York.