

SEGMENTACIÓN DE CURVAS DE LACTANCIA DE BOVINOS, MEDIANTE CLUSTER ANÁLISIS Y ANÁLISIS DISCRIMINANTE LINEAL, APLICADO AL PRIMER TERCIO DE LACTANCIA

SEGMENTATION OF BOVINE LACTATION CURVES, DURING THE FIRST THIRD OF LACTATION VIA CLUSTER AND LINEAL DISCRIMINANT ANALYSIS

Claudio Sebastián Cárdenas Mansilla

Ingeniero Agrónomo. MBA. Mg.

Universidad de los Lagos Sede Castro, Chiloe. ccard007@gmail.com, ccard007@hotmail.com

ABSTRACT

Key Words: Lactation curve, Cluster Analysis, Discriminant Analysis

This paper presents the results of a segmentation study with Hierarchical Cluster Analysis Ward's algorithm, applied to the data from 465 lactations of dairy cows in the Araucanía region. Only the first third of the lactation was considered with data from measurement intervals of 30 days this, obtaining an analysis matrix of 465 rows (lactations) x 5 columns (checks every 30 days). Validation and evaluation of the cluster solution, based on the clustering of cases (individuals), was performed by linear discriminant analysis. A segment was obtained that allowed the dairy cows to be grouped into 4 productive groups. The groups differed in production volume, but mainly in the type of production in terms of the patterns of production, with normal and atypical production curves. The implementation of the multivariate statistical tools, to analyse of groups and patterns of bovine lactation curves, is very relevant and highly efficient, validated in this study by linear discriminant analysis.

RESUMEN

Palabras Claves: Curva de lactancia, Cluster Análisis, Análisis Discriminante

Este trabajo presenta los resultados de un estudio de segmentación, mediante Análisis Cluster Jerárquico a través del algoritmo de Ward, de 465 lactancias de vacas lecheras en Región de la Araucanía, estudiando solo lactancias hasta el primer tercio de producción, a través de controles mensuales a intervalos de medición de 30 días, obteniendo de esta forma una matriz de análisis de 465 filas (lactancias) x 5 columnas (controles cada 30 días). La validación y evaluación de la solución cluster, en función de la conglomeración de casos (individuos), se realizó mediante Análisis Discriminante Lineal. A modo de conclusión, se obtuvo una segmentación que permitió la agrupación de vacas lecheras en 4 grupos productivos que se diferencian en volumen productivos pero fundamentalmente en tipología de producción en lo referente a forma del patrón de cada grupo productivo, es decir, agrupando las lactancias en clusters de curvas de lactancias de producción normal y curvas de producción atípicas. En tal sentido la aplicación de esta herramienta estadística Multivariante, en el análisis de grupos y patrones de curvas de lactancia bovina, resulta muy pertinente y altamente eficiente, validada en este estudio mediante análisis discriminante lineal.

INTRODUCCIÓN.

El estudio de las propiedades matemáticas de la curva de lactancia proporciona información resumida acerca de la producción de ganado lechero, que es útil en la toma de decisiones de gestión, por ejemplo, la vigilancia de la salud y la alimentación individual (Silvestre *et al.*, 2006).

Los modelos matemáticos empíricos de curvas de lactancia son funciones regulares y $= f(t)$, definida por los valores positivos de la producción diaria de leche (y) y el tiempo de parto (t), utilizado en la industria de cría de ganado lechero y de gestión. Estos modelos representan una herramienta de investigación esencial para el desarrollo y la validación de los modelos mecanicistas, destinada a explicar las principales características de la estructura de producción de leche en términos de la biología conocida de la glándula mamaria durante el embarazo y lactancia (Macciotta *et al.*, 2005).

La necesidad de aumentar la eficiencia del sistema de producción de leche hace conveniente conglomerar los animales en grupos productivos, considerando que una evaluación importante en cualquier sistema de producción, es la detección de unidades productivas que presenten bajo rendimiento o anomalías, que afectan la eficiencia global del sistema. Esto en la producción lechera, implica la identificación oportuna de las curvas de lactancia que presentan comportamientos que se escapan a las tendencias consideradas normales (Cárdenas, 2008a). En la actualidad, la atención se centra particularmente en la capacidad de la vaca para mantener la producción después del pico de la persistencia (la lactancia), la importancia económica de la que se encuentra en sus relaciones con el estado de salud, la eficiencia de reproducción, y alimentación (Macciotta *et al.*, 2004)

En las palabras de Roure (1996), para manejar eficientemente un sistema de producción de leche, es necesario disponer de antecedentes que permitan la evaluación, y el análisis permanente, puesto que es necesario evaluar la situación inicial como el avance a través del tiempo. La primera etapa en todo diseño de reconocimiento, consiste en el establecimiento

de las clases: en lo que se podría denominar como la definición del universo de trabajo del sistema, lo que en la mayoría de los casos es directa y trivial (Maravall, 1993). No obstante, puede ocurrir que las clases sean desconocidas *a priori*, situación que se presenta en ciertos campos de la biología en donde no está clasificado el universo de clases. En tales situaciones, se recurre a técnicas denominadas de conglomeración o clustering (Cárdenas, 2003).

En este sentido la estadística multivariable exploratoria ha sido empleada en muchas áreas de la ciencia aplicada para clasificar y establecer relaciones de similitud entre unidades de las cuales se han medido una gran cantidad de variables (Smith *et al.*, 2002). El Análisis de Conglomerados o Cluster Análisis agrupa a los individuos u objetos en conglomerados de tal forma que los objetos del mismo conglomerados son más parecidos entre sí que, respecto a los objetos de otros conglomerados, lo que se intenta con esto, es maximizar la homogeneidad dentro de los conglomerados, mientras que a la vez, se maximiza la heterogeneidad entre los agregados (Hair *et al.*, 1999).

El Análisis Discriminante es una técnica estadística que permite estudiar las diferencias entre dos o más grupos de objetos con respecto a varias variables, simultáneamente. Es una técnica de clasificación y asignación de individuos a grupos, a los cuales, se les conoce sus características (Cuesta, 1992).

Este trabajo tiene como objetivo general la aplicación de herramientas estadísticas multivariantes, como el Análisis Clúster Jerárquico, en la segmentación de grupos y patrones de curvas de lactancia bovina, y la validación de la misma mediante Análisis Discriminante Lineal.

MATERIALES Y METODOS

El trabajo se realizó a partir de información de lactancias recopilada en el rebaño lechero de la Estación Experimental Maipo de la Universidad de la Frontera y de los controles lecheros de la actual Estación Experimental de la misma Universidad; el Fundo Maquehue y

además, se contó con la información de control lechero oficial recopilada por la SOFO, en la Región de la Araucanía, La muestra se elaboró en forma selectiva con relación a la tendencia de producción de leche, seleccionando sólo lactancias completas en su primer tercio, las cuales se distribuyen desde el año 1997 hasta el año 2001. Esta muestra estuvo conformada por lactancias controladas mensualmente (cada 30 días aproximadamente), considerando solo 5 producciones por cada animal a intervalos de 30 días, desde el día 8 (prod 0) al día 128 (prod 4) de la lactancia, logrando obtener una muestra de 465 lactancias individuales.

Una vez conformada la matriz de análisis con las 465 entradas (registros individuales) y 5 columnas se procedió a conformar los conglomerados según los datos presentes en la muestra. Esto se realizó utilizando Cluster Jerárquico, a través del algoritmo de Ward, (Peña, 2002).

$$W = \sum_{g=1}^G \sum_{i=1}^{n_g} (\mathbf{x}_{ig} - \bar{\mathbf{x}}_g)(\mathbf{x}_{ig} - \bar{\mathbf{x}}_g)^T$$

Donde:

W = Matriz de suma de cuadrados dentro del grupo

X_{ig} = Valor del dato i en el grupo g

\bar{X}_g = Media del grupo g

G = N° de grupos prefijados

Este método parte de los elementos (observaciones) directamente, en lugar de utilizar la matriz de distancias, y se define una medida global de heterogeneidad de una agrupación de observaciones en grupos (Cárdenas, 2008b). Luego se realizó un análisis al dendrograma obtenido y del porcentaje de cambio del coeficiente de aglomeración, a partir de esto se determinó el número de segmentos presentes. Finalmente se aplicó análisis discriminante lineal para determinar si existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre el

conjunto de variables (5 niveles productivos tomados cada 30, días durante el primer tercio de la lactancia), de los segmentos ya definidos a través de análisis cluster (Hair *et al.*, 1999).

Según Hair *et al.*, (1999) en un determinado problema se halla una combinación lineal de variables independientes, obteniéndose una serie de puntuaciones D_s , para cada objeto en cada grupo. Las puntuaciones D_s son llamadas "puntuaciones, discriminantes" y se calculan de acuerdo a la regla estadística que permite maximizar las varianzas entre los grupos y minimizar la varianza intra-grupo.

Formalmente es Ferran (2001):

$$D_s = B_{s1}X_1 + \dots + B_{sp}X_p + B_{s0}$$

Si la varianza entre-grupo es grande en relación a la varianza intra-grupo, se puede decir que la función separa bien los grupos.

RESULTADOS Y DISCUSIÓN

El proceso de clusterización mediante segmentación jerárquica, utilizando el método de Ward, permitió identificar 4 segmentos de producción, de curvas de lactancias, al primer tercio, presentados en la siguiente figura; (figura 1).

En estos términos son claros los de grupos de individuos segmentados, sin embargo no los hemos identificado, no conocemos sus perfiles, en esta situación la exposición de los perfiles resulta fundamental para identificar el patrón de producción de cada conglomerado de lactancia identificado, observado en la Figura 2.

Evidentemente se puede observar que de los 4 patrones, que varían según su forma y nivel de producción, se tienen tres patrones de producción normal, en términos de semejanza con la Curva de Wood (Wood, 1980), en tres niveles productivos diferentes. Por otra parte se observa un patrón de producción atípica con un cluster de curvas de muy alta producción pero irregulares en el perfil. En función de lo anterior podríamos clasificar las producciones

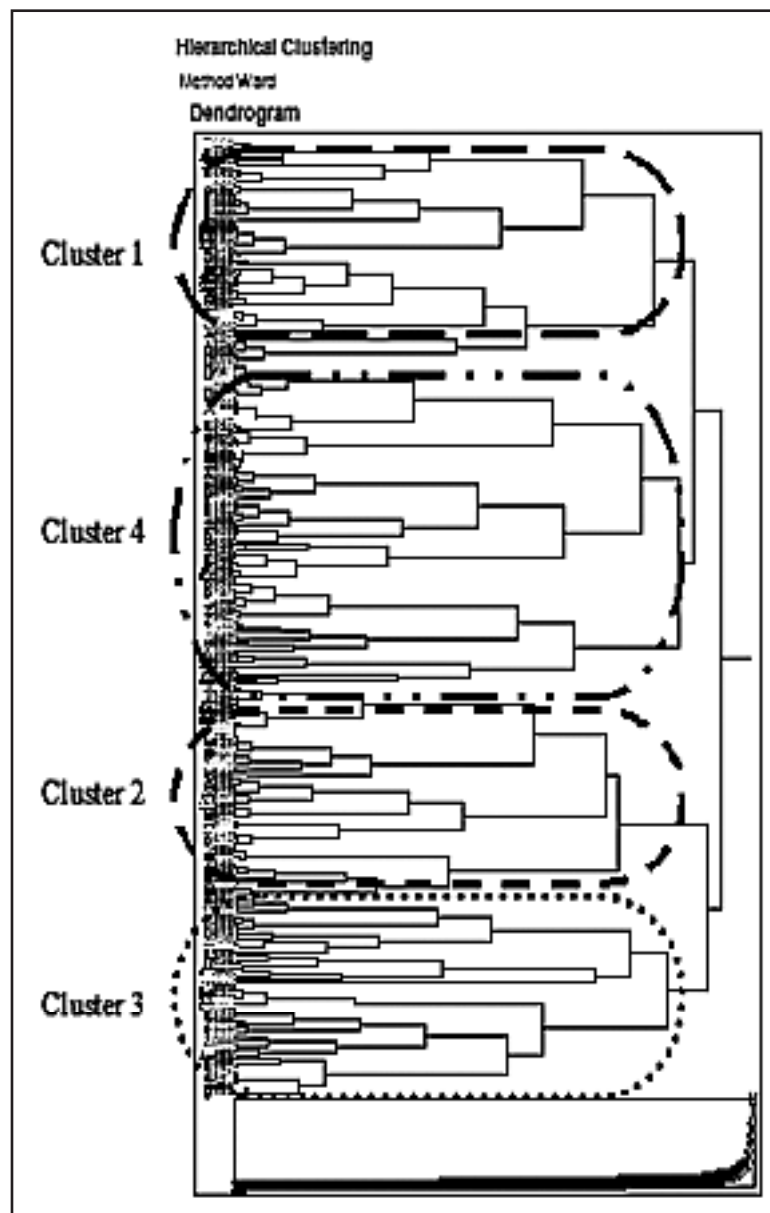


Figura 1. Dendrograma de aglomeración de curvas de latencia.
Figure 1. Dendrogram of agglomeration lactation curves

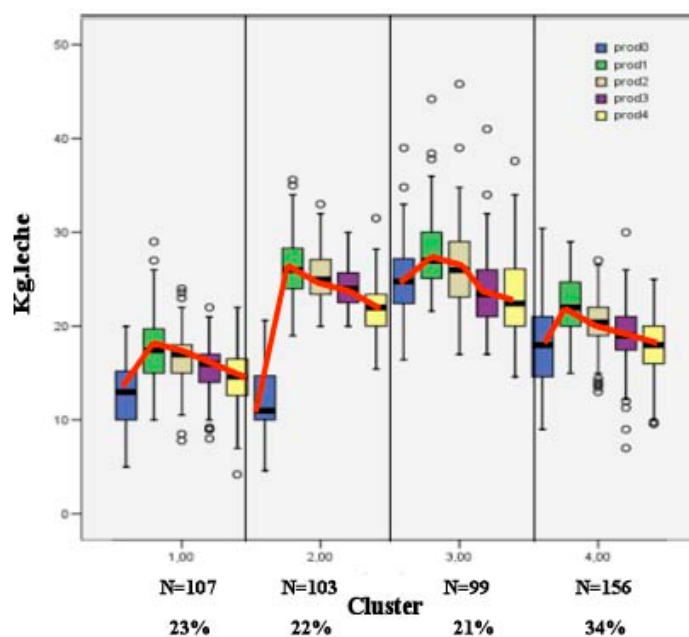


Figura 2 . Perfiles de tipologías identificadas, mediante gráficos de caja y rango.
Figure 2. Profiles of types identified by box plots and rank

de la siguiente manera:

Para validar la segmentación y determinar si existen diferencias estadísticas significativas entre los perfiles de las puntuaciones medias sobre el conjunto de cinco variables, se aplica el Análisis Discriminante Lineal (Hair *et al*, 1999), sobre los cuatro grupos definidos anteriormente, mediante cluster jerárquico. Al aplicar análisis Discriminante, con los fines antes expuestos, tenemos lo siguiente;

Viendo la dispersión de cada cluster alrededor del centroide de grupo, podemos observar que los clusters de *Perfiles de Producción Normal* (Clusters 1, 2 y 4), están más concertados con pequeños solapamientos entre sí, sin embargo el *Perfil de Producción Atípica*, presenta una menor concentración fundamentalmente en la dimensión de Función Discriminante 1, que discrimina en función de las variables, prod1, prod2, prod3 y prod4 (peak y persistencia), ver

Cuadro 1. Parámetros de Clústers identificados mediante aglomeración jerárquica.

Table 1. Parameters of clusters identified by hierarchical agglomeration.

Cluster	Nominación	Producción Mediana Inicial (día kg/leche/día)	Producción Mediana al Peak kg/leche/día
Cluster 1	Perfil de baja producción normal	13,0	17,4
Cluster 2	Perfil de alta producción normal	11,0	26,0
Cluster 3	Perfil de producción atípico	24,8	27,0
Cluster 4	Perfil de producción media normal	18,0	22,0

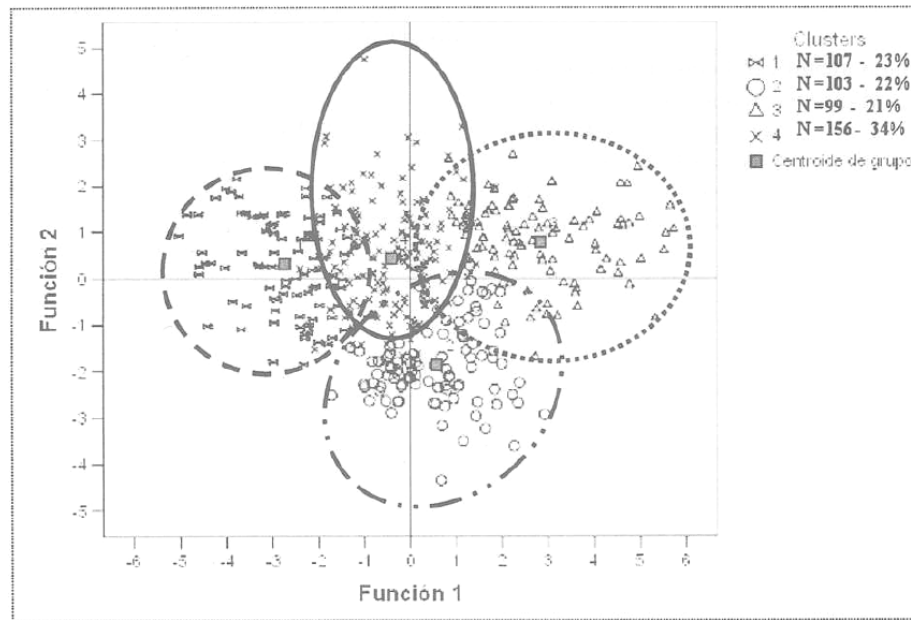


Figura 3 . Diagrama de dispersión de todos los casos en el plano definido por las dos primeras funciones discriminantes.

Figure 3. Scatterplot of all cases in the plane defined by the first two discriminant functions.

cuadro 2, según el cual, estas lactancias tienen una mayor dispersión en el periodo desde el peak al final del primer tercio de producción (128 días). Finalizando el análisis de la figura 3, muestra el diagrama de dispersión de todos los casos en el plano definido por las dos primeras funciones discriminantes, en el cual, se observa que los grupos presentan separaciones territoriales claramente definidas sin solapamientos significativos, de esta forma, la mejor separación de los individuos de cada

uno de los grupos obtenidos mediante análisis de clúster jerárquico, es lograda de buena manera por las 2 primeras funciones discriminantes (ver, cuadro 2), pudiendo comprobar que; el valor teórico obtenido puede clasificar con alta certeza los vectores presentados, en el conglomerado al cual debiera pertenecer, determinando de esta manera que existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre el conjunto de vectores (controles) analizados.

Cuadro 2. Matriz Estructura. Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas - Variables ordenadas por el tamaño de la correlación con la función.

Table 2. Matrix Structure. Combined within-group correlations between discriminating variables and canonical discriminant functions typified - Variables ordered according to size of correlation with function.

Cluster	Función Discriminante		
	1	2	3
prod1	0,59790818 *	-0,33115919	0,46498783
prod2	0,59691449 *	-0,44362127	-0,58242604
prod3	0,54007797 *	-0,53595755	-0,11429482
prod4	0,49725373 *	-0,341487	0,02516907
prod0	0,54400826	0,80196304 *	-0,09385801

* Mayor correlación absoluta entre cada variable y cualquier función discriminante.

CONCLUSIONES.

Se pudieron identificar 4 clusters productivos diferentes, al primer tercio de la lactancia, donde el 79% de las lactancias estudiadas están clasificadas dentro de un perfil de producción con forma normal y un 21% de las lactancias se clasifican dentro de un perfil de producción atípica.

Se pudo observar que lactancias que se inician con una alta producción, sobre 20 kg/leche/día (al día 8), puede hacer prever un patrón de su curva de producción irregular y generalmente atípico. Por el contrario lactancias que se inician con una baja producción, por debajo de 18 kg/leche/día (al día 8), puede hacer prever un patrón de su curva regular en su forma normal.

Según Cárdenas (2003), la segmentación estadística realizada mediante Cluster Jerárquico, facilita la agrupación de vacas lecheras en diferentes clases productivas de cuyos centroides se pueden generar patrones de lactancias. Esto se concluyó mediante una muestra de 188 lactancias, en este estudio utilizando una muestra más amplia de 465 lactancias, se pudo obtener una segmentación que permitió corroborar algo similar; la agrupación de vacas lecheras en 4 grupos productivos que se diferencian en volumen productivos pero fundamentalmente en tipología de producción en lo referente a forma del patrón de cada grupo productivo, es decir, agrupando las lactancias en clusters de curvas de lactancias de producción normal y curvas de producción atípicas. En tal sentido la aplicación de esta herramienta estadística multivariante, en el análisis de grupos y patrones de curvas de lactancia bovina, resulta muy pertinente y altamente eficiente, validada en este estudio mediante análisis discriminante lineal.

AGRADECIMIENTOS

Quiero agradecer de manera especial al Prof.: Horacio Miranda V. y al Prof.: Sergio Hazard T., del Departamento de Producción Agropecuaria de la Universidad de la Frontera, por su incentivo constante, durante mis años de alumno en la universidad, y por haberme permitido acceder

a la información con la, cual se realizan mis investigaciones.

BIBLIOGRAFIA

- CÁRDENAS, C.S. 2003. Reconocimiento de patrones y clasificación de curvas de lactancia mediante redes neuronales y análisis discriminante lineal aplicados al primer tercio de la lactancia a controles de vacas lecheras de la IX región. Tesis de Ingeniero Agrónomo. Universidad de la Frontera. Temuco, Chile. 67 p.
- CÁRDENAS, C. S. 2008 a. Reconocimiento de patrones de curvas de lactancia mediante red neuronal y análisis discriminante lineal, al primer tercio de lactancia a controles de vacas lecheras de la IX región. Agro sur, Mayo 2008, vol.36, no.1, p. 43-48.
- CÁRDENAS, C. S. 2008 b. Identificación de tipologías de actitud hacia las matemáticas en estudiantes de séptimo y octavo grados de educación primaria. Perfiles educativos, 30: p. 94-108.
- CUESTA, M. 1992. Análisis Discriminante. In: Vallejo, G. (e.d.) Análisis Multivariantes Aplicados A las Ciencias del Comportamentales. Universidad de Oviedo. Oviedo, España. 286 p.
- FERRAN, M. 2001. SPSS Para Windows, Análisis Estadístico. Primera Edición en español. Madrid, España, McGRAW-HILL/INTERAMERICANA DE ESPAÑA, S.A.U. 421 p.
- HAIR,S.F., ANDERSON, R.E.; TATHAM , R.L.; BLACK., W.C. 1999. Madrid, España. Análisis Multivariante. Quinta edición en español. Ed. Prentice Hall. Inc. 832 p.
- PEÑA, D. 2002. Análisis de datos Multivariantes. Primera Edición en español. Madrid, España. Editorial MacGrawHill. 539 p.
- MACCIOTTA, N.; VICARIO, D.; CORRADO DI MAURO; CAPPIO-BORLINO, A. 2004. A Multivariate Approach to Modeling Shapes of Individual Lactation Curves in Cattle. American Dairy Science Association. J. Dairy Sci. 87: 1092-1098
- MACCIOTTA, N.; VICARIO, D. ; BORLINO. A. 2005. Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. American Dairy Science Association. J. Dairy Sci. 88: 1178-1191.
- MARAVALL, D. 1993. Reconocimiento de Formas y Visión Artificial. Primera Edición. RA- MA Editorial, Madrid, España. 433 p.
- ROURE, S. 1996. Desarrollo de un método de control de predicción y de un coeficiente de discrepancia

- de producción máxima esperada , en curva de lactancia de ganado bovino lechero. Tesis de Ingeniero Agrónomo. Universidad de la Frontera. Temuco, Chile. 89 p.
- SMIHT, R.; MOREIRA, V.; LATRILLE, L. 2002. Caracterización de Sistemas Lecheros en la X Región de Chile Mediante Análisis Multivariable. *Agric. Téc.* 62: 375-395.
- SILVESTRE, A. PETIM-BATISTA, F. and COLAÇO, J. 2006. The Accuracy of Seven Mathematical Functions in Modeling Dairy Cattle Lactation Curves Based on Test-Day Records From Varying Sample Schemes. *American Dairy Science Association. J. Dairy Sci.* 89: 1813-1821.
- WOOD, P. 1980. Relationships between size, live – weight change and milk production characters in early lactation in dairy cattle. *Animal Production* 31: 143–151.