

Multi-trait selection in Holstein cows

Selección multicarácter en vacas Holstein.

A. Hernández and Raquel Elena Ponce de León

Instituto de Ciencia Animal, Apartado Postal 24, San José de Las Lajas, Mayabeque, Cuba.

Email: arelishdez@ica.co.cu

A. Hernández: <https://orcid.org/0000-0001-5123-5123>

Raquel Elena Ponce de León: <https://orcid.org/0000-0001-8632-7036>

The phenotypic and genealogical data of 1 571 Holstein cows located in three livestock enterprises during the years 1984 to 2016 were used. The purpose of this study was to carry out the multi-trait selection of dairy production, reproduction and longevity traits by means of the preparation of selection indexes (SI) through the analysis of main components (MC). Correlations and breeding values (BV) were estimated for the traits: cumulative milk production up to 305 days (BV_{L305}), duration of lactation (BV_{DL}), age at first parturition (BV_{API}), gestation parturition interval (BV_{GPI}), accumulated milk per life (BV_{TML}) and productive life (B_{VP}) by means of a multi-trait animal model. The SPSS statistical package was used to perform the MC analysis, for which the BV were standardized, and the Kaiser criterion was used to select the MC that explains the greatest genetic variation. The genetic correlations between L305, TML and DL showed mean values (0.55, 0.47 and 0.27), and between GPI and PL of 0.29. The first two main components (MC_1 , MC_2) were those that obtained the Kaiser criterion and explained 53.7 % of the total variance of the BV. Linear correlations between BVs with each main component showed that L305, DL and TML were related with the MC_1 , and GPI and PL to MC_2 . It is concluded that in Holstein cows it is possible to perform multi-trait selection by constructing selection indexes based on the first two MCs, since they showed considerable genetic variation.

Key words: *main component analysis, dairy cattle, selection indexes*

Genetic evaluation is a process that allows obtaining the breeding value of the animals for one or more characteristics and in that way selecting as reproducers those with the greatest genetic merit (Ruales *et al.* 2007). The mixed model methodology allows obtaining the BLUP (Best Linear Unbiased Prediction) or MPLI (Best Linear Unbiased Prediction) of the breeding value of the animals evaluated based on their own records and their relatives records (Kennedy *et al.* 1988).

The traditional selection indexes (Smith 1936 and Hazel 1943) are used to select animals by linear combinations of breeding values with economic weights. However, it is sometimes difficult to obtain the economic weights (the information is not available), in addition to being subject to constant market variations.

An alternative was to develop an index in which each trait had a weight to obtain the gain desired by the farmer without considering the relative importance of the traits in an economic sense (Rouvier 1969

Se utilizaron los datos fenotípicos y genealógicos de 1 571 vacas Holstein ubicadas en tres empresas ganaderas durante los años 1984 a 2016. El propósito del estudio fue realizar la selección multicarácter de rasgos de producción lechera, reproducción y longevidad mediante la confección de índices de selección (IS) a través del análisis de componentes principales (CP). Se estimaron las correlaciones y valores genéticos (VG) para los rasgos: producción de leche acumulada hasta los 305 días (VG_{L305}), duración de la lactancia (VG_{DL}), edad al primer parto (V_{GEP1}), intervalo parto gestación (VG_{IPG}), leche acumulada por vida (VG_{LTV}) y vida productiva (VG_{VP}) mediante un modelo animal multicarácter. Se utilizó el paquete estadístico SPSS para realizar el análisis de CP para lo cual estandarizaron los VG, y se usó el criterio de Kaiser para seleccionar el CP que explica la mayor variación genética. Las correlaciones genéticas entre L305, LTV y DL presentaron valores medios (0.55, 0.47 y 0.27), y entre IPG y VP de 0.29. Los dos primeros componentes principales (CP_1 , CP_2) fueron los que obtuvieron el criterio de Kaiser y explicaron el 53.7 % de la varianza total de los VG. Las correlaciones lineales entre los VG con cada componente principal mostraron que L305, DL y LTV estuvieron relacionados con el CP_1 , y IPG y VP con el CP_2 . Se concluye que en vacas Holstein es posible realizar la selección multicarácter mediante la construcción de índices de selección basados en los dos primeros CP, pues los mismos mostraron considerable variación genética.

Palabras clave: *Análisis de componentes principales, ganado lechero, índices de selección*

La evaluación genética es un proceso que permite obtener el valor genético de los animales para una o más características y así seleccionar como reproductores aquellos con mayor mérito genético (Ruales *et al.* 2007). La metodología de modelos mixtos permite obtener el BLUP (Best Linear Unbiased Prediction) o MPLI (Mejor Predicción Lineal Inesgada) del valor genético de los animales evaluados en función de sus propios registros y registros de sus parientes (Kennedy *et al.* 1988).

Los índices de selección tradicionales (Smith 1936 y Hazel 1943) son utilizados para seleccionar los animales por combinaciones lineales de los valores genéticos con los pesos económicos. Sin embargo, en ocasiones resulta difícil obtener los pesos económicos (no se dispone de la información), además de estar sujetos a las constantes variaciones del mercado.

Una alternativa fue desarrollar un índice en el cual cada rasgo tuviera un peso para obtener la ganancia deseada por el productor sin considerar la importancia relativa de los rasgos en sentido económico (Rouvier 1969 y Yamada *et al.* 1975). Pesek y Baker (1969)

and Yamada *et al.* 1975). Pesek and Baker (1969) suggest that the farmer should specify the amount of improvement required for each characteristic to achieve the desired gain. Thus, the desired gain is incorporated with the estimates of the genetic parameters in the selection index.

Recently, several countries such as Brazil and India have incorporated the main component analysis (MC) in the construction of selection indexes in dairy cattle (Bignardi *et al.* 2012 and Khan *et al.* 2013). Authors such as Buzanskas *et al.* (2013) stated that the use of MC is a methodology to build linear combinations between the breeding values of the available traits in a database, taking into account the eigenvalues of the main component and the eigenvectors of the traits in each main component, which are variability measures. The great advantage of this procedure is that the eigenvectors are orthogonal, that is, they are not correlated and therefore they can be added together, in this way the selection based on the index maximizes the genetic merit for the included characteristics.

This study was carried out with the purpose of carrying out the multi-trait selection of milk production, reproduction and longevity traits in Holstein cows by means of the preparation of selection indexes (SI), through the analysis of main components.

Materials and Methods

Description of the database. The information on milk production and reproduction, which was registered in the Sistema de Control Pecuario (SISCOP), of Holstein cows with parturitions between 1984 and 2016 was used. These cows were located in three livestock enterprises (Empresa Pecuaria Genética de Matanzas in Matanzas province, Los Naranjos in Mayabeque and Camilo Cienfuegos in Pinar del Río) in the western region of the Republic of Cuba.

The records of live animals were used as well as of those that had caused leave. From the milk production data file (live + pouled) were obtained the traits: Cumulative milk production up to 305 days (L305), duration of lactation (DL) and age at first parturition (AP1). While, from the reproduction data file (live + pouled) the gestation parturition interval (GPI) was calculated. From the file of pouled of milk production, accumulated milk for life (TML) was calculated; as well as the longevity trait called productive life (PL) determined as the months from the first to the last parturition.

Creation of contemporaries groups. A general linear model (GLM) was used using the statistical package SAS (2013) version 9.3 to define the significant fixed effects ($P < 0.01$) to be included in contemporaries group (CG). The herd-year- parturition season combination was considered as CG and those groups made up of less than 3 animals were eliminated. Two parturition seasons were determined: the rainy (from May to October) and

sugieren que el productor debe especificar la cantidad de mejora requerida para cada característica para alcanzar la ganancia deseada. Así, la ganancia deseada es incorporada con los estimados de los parámetros genéticos en el índice de selección.

Recientemente varios países como Brasil y la India incorporaron el análisis de componentes principales (CP) en la construcción de índices de selección en el ganado lechero (Bignardi *et al.* 2012 y Khan *et al.* 2013). Autores como Buzanskas *et al.* (2013) expusieron que el uso de los CP es una metodología para construir combinaciones lineales entre los valores genéticos de los rasgos disponibles en una base de datos, tomando en consideración los valores propios del componente principal y los vectores propios de los rasgos en cada componente principal, los cuales son medidas de variabilidad. La gran ventaja de este procedimiento es que los vectores propios son ortogonales, es decir no están correlacionados y por tanto que se pueden sumar, de esta forma la selección basada en índice maximiza el mérito genético para las características incluidas.

El presente trabajo se realizó con el propósito de realizar la selección multicarácter de rasgos de producción lechera, reproducción y longevidad en vacas Holstein mediante la confección de índices de selección (IS), a través del análisis de componentes principales.

Materiales y Métodos

Descripción de la base de datos. Se utilizó la información de la producción lechera y la reproducción, que se registró en el Sistema de Control Pecuario (SISCOP), de las vacas Holstein con partos entre 1984 al 2016. Estas vacas se ubicaron en tres empresas ganaderas (Empresa Pecuaria Genética de Matanzas en la provincia de Matanzas, Los Naranjos en Mayabeque y Camilo Cienfuegos de Pinar del Río) de la región occidental de la República de Cuba.

Se utilizaron los registros de los animales vivos y también los de aquellos que habían causado baja. A partir del fichero de datos de la producción lechera (vivas + bajas) se obtuvieron los rasgos: Producción de leche acumulada hasta los 305 días (L305), duración de la lactancia (DL) y edad al primer parto (EP1). Mientras que, del fichero de datos de la reproducción (vivas + bajas) se calculó el intervalo parto gestación (IPG). Del fichero de bajas de la producción lechera se calculó la leche acumulada por vida (LTV); así como el rasgo de longevidad denominado vida productiva (VP) determinada como los meses del primer al último parto.

Creación de grupo de contemporáneos. Se utilizó un modelo lineal general (GLM) mediante el paquete estadístico SAS (2013) versión 9.3 para definir los efectos fijos significativos ($P < 0.01$) a incluir en el grupo de contemporáneos (GC). Se consideró la combinación de rebaño-año-época de parto, como GC y se eliminaron aquellos grupos conformados por menos de 3 animales. Se determinaron dos épocas de parto: la lluviosa (de mayo a octubre) y la poco lluviosa (de noviembre a abril).

the dry (from November to April).

Subsequently, all the files were joined, leaving only those cows that had BV information for all the studied traits. The final file showed a total of 1,571 Holstein cows. The pedigree file was made up by a total of 153,963 individuals. The pedigree information reached the grandparents, through the maternal and paternal lines.

Correlation estimation. The ASREML program (Gilmour *et al.* 2003) was used to estimate the genetic and environmental correlations and the breeding values (BV) using the following multi-trait animal model:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} x_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & x_6 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} + \begin{bmatrix} z_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & z_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & z_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & z_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & z_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & z_6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

Where:

y_i = vector of observations for the i -th trait

b_i = vector of fixed effects (herd-year-parturition season and parturition age as linear and quadratic covariates) for the i -th trait

a_i = vector of the random effects of the animal for the i -th trait

e_i = vector of random residual effects for the i -th trait

x_{ij} y Z_{ij} = design matrices relating data to fixed and random effects, respectively

Main components analysis. The main components analysis was carried out using the SPSS (2002) statistical package version 11.5. This analysis was performed as a way to condense or summarize the information contained in several original variables (in this case the BVs) into a smaller group of new composite dimensions or variants called main components, with minimal loss of information, and to explore the relations between the BVs obtained from the single-trait analysis BV_{L305}, BV_{DL}, BV_{TML}, BV_{API}, BV_{GPI}, BV_{VPL}, to explain the data structure (Hair *et al.* 2009).

Due to differences in units of measurement, the BVs for all traits were standardized. The Kaiser (1960) criterion was used to select the main component that explains the highest genetic variation in the data. This criterion takes into consideration only those main components with eigenvalues above unity. The eigenvalue of a main component is associated with the variance of all the traits included in the main component. Each eigenvalue is associated with a unit vector called an eigenvector (Rencher 2002). The eigenvectors have the strength and direction of the variance of each trait with the main component. In this study, a variable correlation matrix was used to obtain eigenvalues.

The first main component explains the largest

Posteriormente se unieron todos los ficheros dejando solo aquellas vacas que tuvieron información de los VG para todos los rasgos estudiados. El fichero final presentó un total de 1 571 vacas Holstein. El fichero de pedigrí quedó conformado por un total de 153 963 individuos. La información del pedigrí alcanzó hasta los abuelos, por la línea materna y por la paterna.

Estimación de correlaciones. Se empleó el programa ASREML (Gilmour *et al.* 2003) para estimar las correlaciones genéticas y ambientales, y los valores genéticos (VG) mediante el siguiente modelo animal multicarácter:

$$\begin{bmatrix} z_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & z_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & z_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & z_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & z_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & z_6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix}$$

Donde:

y_i = vector de las observaciones para el i -ésimo rasgo

b_i = vector de los efectos fijos (rebaño-año-época de parto y la edad al parto como covariable lineal y cuadrática) para el i -ésimo rasgo

a_i = vector de los efectos aleatorios del animal para el i -ésimo rasgo

e_i = vector de los efectos residuales aleatorios para el i -ésimo rasgo

x_{ij} y Z_{ij} = matrices de diseño que relacionan los datos con los efectos fijos y aleatorios, respectivamente

Análisis de componentes principales. El análisis de componentes principales se realizó mediante el paquete estadístico SPSS (2002) versión 11.5. Este análisis se realizó como una vía para condensar o resumir la información contenida en varias variables originales (en este caso los VG) en un conjunto más pequeño de nuevas dimensiones compuestas o variantes llamadas componentes principales, con una pérdida mínima de información, y para explorar las relaciones entre los VG obtenidos del análisis unicarácter VG_{L305}, VG_{DL}, VG_{LTV}, VG_{GPI}, VG_{IPG}, VG_{VPL} para explicar la estructura de los datos (Hair *et al.* 2009).

Debido a las diferencias en las unidades de medida, los VG para todos los rasgos se estandarizaron. Se utilizó el criterio de Kaiser (1960) para seleccionar el componente principal que explica la mayor variación genética de los datos. Este criterio toma en consideración solo aquellos componentes principales con valores propios por encima de la unidad. El valor propio de un componente principal está asociado con la varianza de todos los rasgos incluidos en el componente principal. Cada valor propio está asociado con un vector unitario llamado vector propio (Rencher 2002). Los vectores propios representan la fortaleza y dirección de la varianza de cada rasgo con el componente principal. En el presente trabajo se utilizó una matriz de correlación de variables para obtener valores propios.

percentage of the total variance of the BVs. The second explains the second largest percentage, until all the variance in the database is explained.

Each main component can generate a new value called the main component score, which is the sum of the standardized BVs of the weight of each trait by its respective standardized score coefficient (SSC). In this way, the main component can be used as an index to evaluate animals, for multiple traits. The standardized scoring coefficients of each BV in each main component were obtained using the following formula:

$$SSC_{ij} = \frac{\text{eigen vector } ij}{\sqrt{\text{eigenvalue}_j}}$$

Where:

SSC_{ij} = Standardized scoring coefficients of the BVs of each trait in the j-th main component

The main component (index) score was calculated as:

$$\sum_{i=1}^l SSC_{ij} BV_{il}$$

Where :

MC_{ji} = Main component score (index score) for the l-th animal in the j-th main component

SSC_{ij} = Standardized scoring coefficients of the BVs of the i-th trait in the j-th main component

BV_{il} = Standardized estimated breeding value of the i-th trait in the l-th animal

Results and Discussion

Correlation estimation. The estimates of genetic and environmental correlations in Holstein are showed in table 1. The genetic correlations between cumulative milk up to 305 days, milk production per life, and duration of lactation were moderate, so that selection for milk production will also improve, in some way, the accumulated milk per life and the duration of lactation. While, the genetic correlation of accumulated milk up to 305 days with the rest of traits was low, this implies that these traits will change almost independently of the milk production.

The previous results show that the Holstein cows with the highest accumulated productions up to 305

El primer componente principal explica el más amplio porcentaje de la varianza total de los VG. El segundo explica el segundo porcentaje más amplio, hasta que toda la varianza de la base de datos sea explicada.

Cada componente principal puede generar un nuevo valor denominado puntuación del componente principal, el cual es la suma de los VG estandarizados del peso de cada rasgo por su respectivo coeficiente de puntuación estandarizado (SSC). De esta forma, el componente principal puede ser usado como un índice para evaluar animales, para múltiples rasgos. Los coeficientes de puntuación estandarizados de cada VG en cada componente principal se obtuvieron utilizando la siguiente formula:

$$SSC_{ij} = \frac{\text{vector propio } ij}{\sqrt{\text{valor propio}_j}}$$

Donde:

SSC_{ij} = Coeficientes de puntuación estandarizados de los VG de cada rasgo en el j-ésimo componente principal

La puntuación del componente principal (índice) se calculó como:

$$\sum_{i=1}^l SSC_{ij} VG_{il}$$

Donde:

CP_{ji} = Puntuación del componente principal (puntuación del índice) para el l-ésimo animal en el j-ésimo componente principal

SSC_{ij} = Coeficientes de puntuación estandarizados de los VG del i-ésimo rasgo en el j-ésimo componente principal

VG_{il} = Valor genético estimado estandarizado del i-ésimo rasgo en el l-ésimo animal

Resultados y Discusión

Estimación de correlaciones. Los estimados de las correlaciones genéticas y ambientales en el Holstein se presentan en la tabla 1. Las correlaciones genéticas entre la leche acumulada hasta 305 días, la producción de leche por vida y la duración de la lactancia fueron moderadas por lo que la selección por producción de leche también mejorará, de alguna forma, la leche acumulada por vida y la duración de la lactancia. Mientras que, la correlación genética de la leche acumulada hasta 305 días con el resto de los rasgos fue baja lo cual implica que estos rasgos cambiarán casi

Table 1. Estimates of genetic (above the diagonal) and environmental (below the diagonal) correlations and standard errors (±) in traits of milk production, reproduction and longevity of Holstein cows.

Traits	L305	DL	TML	AP1	GPI	PL
L305		0.47±0.01	0.55±0.01	-0.20±0.01	0.09±0.01	0.12±0.01
DL	0.35±0.01		0.27±0.02	-0.06±0.01	0.20±0.01	0.21±0.02
TML	0.36±0.01	0.24±0.01		-0.29±0.01	-0.03±0.01	0.05±0.02
AP1	0.01±0.01	0.02±0.01	-0.09±0.01		-0.09±0.02	-0.11±0.02
GPI	0.07±0.01	0.17±0.01	-0.10±0.01	-0.03±0.01		0.29±0.01
PL	0.10±0.01	0.10±0.01	0.75±0.01	-0.13±0.01	0.05±0.01	

L305: accumulative milk production up to 305 days, DL: duration of lactation, TML: accumulative milk per life. AP1: age at first parturition, GPI: gestation parturition interval, and PL: productive life.

days were not the longest-lived. These corroborate previous studies carried out in our country in the same breed (Ponce de León and Guzmán 1993); as well as in the Siboney de Cuba breed (Ponce de León *et al.* 2002) where antagonistic values were found in the genetic correlations between milk in first lactation and longevity (-0.60 to -0.64 and from -0.30 to -0.61, respectively).

The previous performance is probably due to the fact that there was no differential treatment for cows with high production. They are disadvantaged in the diet because the amount of concentrate offered is in correspondence with the average of the high group, which is lower than their average milk production, and therefore they do not cover their nutritional requirements. This causes that the cows with high production had reproductive problems (they do not show heat and are not pregnant) so they are the most likely to cause loss.

The genetic correlations of age at first parturition with L305 and TML were moderate and negative, which shows that these traits, to some extent, will change in the opposite direction. Similar results were seen in the Holstein in Ethiopia by Ayalew *et al.* (2017) where the estimates of the genetic correlations between AP1 and L305 were -0.24 ± 0.11 .

Several authors (Ettema and Santos 2004, Ojango *et al.* 2005 and Ruiz *et al.* 2007) reported that the effect of age at first parturition on longevity and yield of the productive life of the animal was maximized at an age of 20 to 36 months in the first lactation. On the other hand, Buzanskas *et al.* (2010) suggest that animals that give birth at a young age are more probable to have better reproductive yield for a long time.

The analysis of multi-trait data, used in this study, is useful in the selection of the best animals for replacement, since in advance it allows knowing the possible effect of the selection for milk production on the reproductive and longevity traits studied. Authors such as Unalan and Cebeci (2004) refer to the importance of these analyzes by providing estimates of reliable and unbiased genetic parameters. On the other hand, the variance-covariance estimates help to evaluate the magnitude of the genetic correlations between the traits that are the objective of selection, which allow establishing a total merit index for the precise evaluation of the genetic merit of the animals in the herd.

Main components analysis. The main component analysis showed that the first two main components (MC₁, MC₂) reached eigenvalues higher than the unit, thus fulfill the Kaiser criterion (table 2). These explained 53.7 % of the total variance of the breeding values. The results of this study correspond to others where it was also shown that with the use of MC analysis the dimensionality of traits can be reduced.

According to the Bignardi *et al.* (2012) studies in Holstein cows, only two MCs were required to summarize the genetic variation of the bulls among the

independientemente de la producción lechera.

Los resultados anteriores indican que las vacas Holstein con mayores producciones acumuladas hasta 305 días no fueron las más longevas. Estos corroboran estudios anteriores realizados en nuestro país en la misma raza (Ponce de León y Guzmán 1993); así como en la raza Siboney de Cuba (Ponce de León *et al.* 2002) donde se encontraron valores antagónicos en las correlaciones genéticas entre la leche en primera lactancia y la longevidad (-0.60 a -0.64 y de -0.30 a -0.61, respectivamente).

El comportamiento anterior probablemente se deba a que no hubo un tratamiento diferencial para las vacas altas productoras. Las mismas están desfavorecidas en la alimentación pues la cantidad de concentrado que se les ofrece está en correspondencia con la media del grupo de alta, que es inferior a su promedio de producción lechera, y por consiguiente no cubren sus requerimientos nutricionales. Esto provoca que las vacas altas productoras presenten problemas reproductivos (no presentan celo y no quedan gestantes) por lo que son las más propensas a causar baja.

Las correlaciones genéticas de la edad al primer parto con L305 y LTV fueron moderadas y negativas lo cual indica que estos rasgos, en cierta medida cambiarán en sentido contrario. Resultados similares se apreciaron en el Holstein en Etiopía por Ayalew *et al.* (2017) donde los estimados de las correlaciones genéticas entre EP1 y L305 fueron de -0.24 ± 0.11 .

Varios autores (Ettema y Santos 2004, Ojango *et al.* 2005 y Ruiz *et al.* 2007) informaron que el efecto de la edad al primer parto en la longevidad y el rendimiento de la vida productiva del animal, era maximizado con una edad de 20 a 36 meses en la primera lactancia. Por otra parte, Buzanskas *et al.* (2010) sugieren que los animales que paren a una edad temprana tienen más probabilidades de tener mejor rendimiento reproductivo durante un tiempo prolongado.

El análisis de datos multicarácter, utilizado en el presente trabajo, es de utilidad en la selección de los mejores animales para el reemplazo, puesto que de manera anticipada permite conocer el posible efecto de la selección por producción de leche en los rasgos de reproducción y longevidad estudiados. Autores como Unalan y Cebeci (2004) se refieren a la importancia de estos análisis al proporcionar estimados de parámetros genéticos confiables e insesgados. Por otra parte, los estimados de varianza-covarianza ayudan a evaluar la magnitud de las correlaciones genéticas entre los rasgos que son objetivo de selección, las cuales permiten establecer un índice de mérito total para la evaluación precisa del mérito genético de los animales en el rebaño.

Análisis de componentes principales. El análisis de componentes principales mostró que los dos primeros componentes principales (CP₁, CP₂) alcanzaron valores propios superiores a la unidad por lo que cumplen con el criterio de Kaiser (tabla 2). Estos explicaron el 53.7 % de la varianza total de los valores genéticos. Los resultados del presente trabajo se corresponden con otros donde

Table 2. Eigenvalues for each main component, total and accumulated proportion of the additive genetic variance, from the first to the sixth main component (MC₁₋₆), in Holstein cows

Components	Eigenvalues		
	Total	% of the variance additive genetic	% accumulated
MC ₁	2.098	34.972	34.972
MC ₂	1.127	18.781	53.753
MC ₃	0.974	16.232	69.985
MC ₄	0.938	15.632	85.617
MC ₅	0.509	8.479	94.096
MC ₆	0.354	5.904	100.000

10 traits evaluated (10 monthly milk productions). In Canchim cows, Buzanskas *et al.* (2013) reported that 73.37 % of the total variance of the breeding values of three reproductive traits and one growth trait was explained by two MCs. Also in Mambí de Cuba cows (3/4Holstein 1/4Cebú), Hernández and Ponce de León (2018) reported that 53.1 % of the total variance of breeding values was explained by the first two MCs, in dairy production, reproduction and longevity traits.

The linear correlations between the BV of the studied traits in Holstein cows with each main component (table 3) showed that the traits related to milk production and the duration of lactation are highly related to MC₁ while GPI and PL were more associated with MC₂.

también se demostró que con el uso del análisis de CP se puede reducir la dimensionalidad de los rasgos.

De acuerdo con los estudios de Bignardi *et al.* (2012) en vacas Holstein, solo se requerían dos CP para resumir la variación genética de los toros entre los 10 rasgos evaluados (10 producciones mensuales de leche). En vacas Canchim, Buzanskas *et al.* (2013) reportaron que el 73.37 % de la varianza total de los valores genéticos de tres rasgos reproductivos y uno de crecimiento era explicada por dos CP. También en vacas Mambí de Cuba (3/4Holstein 1/4Cebú), Hernández y Ponce de León (2018) informaron que el 53.1% de la varianza total de los valores genéticos era explicada por los dos primeros CP, en rasgos de producción lechera, reproducción y longevidad.

Las correlaciones lineales entre los VG de los rasgos estudiados en las vacas Holstein con cada componente principal (tabla 3) mostraron que los rasgos relacionados

Table 3. Linear correlations of the standardized BV with the main component 1 and 2 (MC₁₋₂), in Holstein cows

BV	Components	
	MC ₁	MC ₂
BV _{L305}	0.842	-0.159
BV _{DL}	0.717	0.337
BV _{TML}	0.728	-0.478
BV _{GPI}	0.284	0.564
B _{VPL}	0.255	0.639
BV _{API}	-0.446	0.184

Los pesos de estos índices son coeficientes de puntuación estandarizados (SSC) para cada VG estandarizado en vacas Holstein (tabla 4). El mayor de los valores absolutos del SSC es el de mayor importancia relativa de los VG estandarizados en el componente principal. Esta importancia es explicada por la mayor correlación lineal entre los rasgos con el componente principal (tabla 3).

con la producción de leche y la duración de la lactancia están altamente relacionados con la CP₁ mientras que el IPG y VP estuvieron más asociados con la CP₂.

Los pesos de estos índices son coeficientes de puntuación estandarizados (SSC) para cada VG estandarizado en vacas Holstein (tabla 4). El mayor de los valores absolutos del SSC es el de mayor importancia relativa de los VG estandarizados en el componente principal. Esta importancia es explicada por la mayor correlación lineal entre los rasgos con el componente principal (tabla 3).

Table 4. Standardized scoring coefficients (SSC) of each estimated standardized BV for all traits in each main component (MC₁₋₂), in Holstein cows

BV	Standardized scoring coefficients	
	SSC (MC ₁)	SSC (MC ₂)
BV _{L305}	0.401	-0.141
BV _{DL}	0.342	0.299
BV _{TML}	0.347	-0.424
BV _{GPI}	0.135	0.500
B _{VPL}	0.122	0.567
BV _{APL}	-0.213	0.164

The main component score (index value) for each Holstein animal, in each main component, was calculated as:

$$MC_1 = 0.401 (BV_{L305}) + 0.342 (BV_{DL}) + 0.347 (BV_{TML})$$

$$MC_2 = 0.500 (BV_{GPI}) + 0.567 (B_{VPL})$$

In Holstein cows, the selection for BV_{L305}, BV_{TML} and BV_{DL} by MC₁ could be carried out separately from the selection for BV_{GPI} and B_{VPL} by MC₂ considering the linear correlation between the BV with each main component (table 3). This corresponds to the genetic correlations between the studied traits (table 1) where there was a mean genetic correlation between L305, TML and DL (0.55, 0.47 and 0.27) and between GPI and PL (0.29).

Using main component analysis, animals can be selected based on just two scores generated by MC₁ and MC₂ instead of the six breeding values. According to Buzanskas *et al.* (2013) when using this approach, the animals can be selected in a balanced way since the scores of each main component are linear combinations of all the breeding values of the evaluated traits, and not empirical weights typically used in improvements programs.

The authors cited above stated that the use of MCs is a methodology to build linear combinations between the breeding values of the traits available in a database, taking into account the eigenvalues of the main component and the eigenvectors of the traits in each main component, which are measures of variability. In this way, traits with low heritability estimates, which are rarely taken into consideration in a direct selection process, can be included in the main component.

In the Holstein cows analyzed in this study, the MC₁ can be considered a genetic index of milk production because it favors genetically superior animals for BV_{L305}, BV_{TML} and BV_{DL}; while the MC₂ can be a genetic index related to reproduction and longevity that would consider the VG_{IPG} and VG_{VP}.

Conclusions

The results show that in Holstein cows it is possible

La puntuación del componente principal (valor del índice) para cada animal Holstein, en cada componente principal, fue calculada como:

$$CP_1 = 0.401 (VG_{L305}) + 0.342 (VG_{DL}) + 0.347 (VG_{LTV})$$

$$CP_2 = 0.500 (VG_{IPG}) + 0.567 (VG_{VP})$$

En vacas Holstein, la selección para VG_{L305}, VG_{LTV} y VG_{DL} mediante la CP₁ podría realizarse por separado de la selección para VG_{IPG} y VG_{VP} a través de la CP₂ considerando la correlación lineal entre los VG con cada componente principal (tabla 3). Lo anterior se corresponde con las correlaciones genéticas entre los rasgos estudiados (tabla 1) donde hubo una correlación genética media entre L305, LTV y DL (0.55, 0.47 y 0.27) y entre IPG y VP (0.29).

Con el uso del análisis de componentes principales se pueden seleccionar los animales basados en solo dos puntajes generados por las CP₁ y CP₂ en lugar de los seis valores genéticos. Según Buzanskas *et al.* (2013) cuando se utiliza este enfoque, los animales se pueden seleccionar de una manera equilibrada una vez que los puntajes de cada componente principal son combinaciones lineales de todos los valores genéticos de los rasgos evaluados, y no pesos empíricos típicamente utilizados en programas de mejora.

Los autores citados anteriormente expusieron que el uso de los CP es una metodología para construir combinaciones lineales entre los valores genéticos de los rasgos disponibles en una base de datos, tomando en consideración los valores propios del componente principal y los vectores propios de los rasgos en cada componente principal, los cuales son medidas de variabilidad. De esta forma, rasgos con bajos estimados de heredabilidad, que raramente son tomados en consideración en un proceso de selección directa, pueden incluirse en el componente principal.

En las vacas Holstein analizadas en el presente estudio la CP₁ puede considerarse un índice genético de producción lechera porque favorece a los animales genéticamente superiores para VG_{L305}, VG_{LTV} y VG_{DL}; mientras que el CP₂ puede ser un índice genético relacionado a la reproducción y la longevidad que consideraría los VG_{IPG} y VG_{VP}.

Conclusiones

Los resultados indican que en vacas Holstein es posible

to perform multi-trait selection by constructing selection indexes based on the first two MCs, since they showed considerable genetic variation.

realizar la selección multicaácter mediante la construcción de índices de selección basados en los dos primeros CP, pues los mismos mostraron considerable variación genética.

References

- Ayalew, W., Aliy, M. & Enyew, N. 2017. "Estimation of genetic parameters of the productive and reproductive traits in Ethiopian Holstein using multi-trait models". *Asian-Australasian Journal of Animal Science*, 30(11): 1550-1556, ISSN: 1011-2367, DOI: <https://doi.org/10.5713/ajas.17.0198>.
- Bignardi, A.B., El Faro, L., Rosa, G.J.M., Cardoso, V.L., Machado, P.F. & Albuquerque, L.G. 2012. "Short communication: Principal components and factor analytic models for test-day milk yield in Brazilian Holstein cattle". *Journal of Dairy Science*, 95(4): 2157-2164, ISSN: 0022-0302, DOI: <https://doi.org/10.3168/jds.2011-4494>.
- Buzanskas, M.E., Grossi, D.A., Baldi, F., Barrozo, D., Silva, L.O.C., Torres Jr, R.A.A., Munari, D.P. & Alencar, M.M. 2010. "Genetic associations between stayability and reproductive and growth traits in Canchim beef cattle". *Livestock Science*, 132(1-3): 107-112, ISSN: 1871-1413, DOI: <https://doi.org/10.1016/j.livsci.2010.05.008>.
- Buzanskas, M.E., Savegnago, R.P., Grossi, D.A., Venturini, G.C., Queiroz, S.A., Silva, L.O., Junior, R.A., Munari, D.P. & Alencar, M.M. 2013. "Genetic parameter estimates and principal components analysis of breeding values of reproduction and growth traits in female Canchim cattle". *Reproduction, Fertility and Development*, 25(5): 775-781, ISSN: 1448-5990, DOI: <https://doi.org/10.1071/RD12132>.
- Ettema, J.F. & Santos, J.P. 2004. "Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms". *Journal of Dairy Science*, 87(8): 2730-2742, ISSN: 0022-0302, DOI: [https://doi.org/10.3168/jds.S0022-0302\(04\)73400-1](https://doi.org/10.3168/jds.S0022-0302(04)73400-1).
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. & Thompson, R. 2003. *ASReml Statistical Software (Release 1.10). User Guide*. Unit of Rothamsted Research, Harpenden, U.K.
- Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. 2009. *Multivariate data analysis*. 7th Ed. Ed. Prentice Hall. Upper SadDL River, Bergen, Nueva Jersey, USA, p. 761.
- Hazel, L.N. 1943. "The genetic basis for constructing selection indexes". *Genetics*, 28(6): 476-490, ISSN: 1943-2631.
- Hernández, A. & Ponce de León, R. 2018. "Índices de selección en ganado Mambí de Cuba mediante el análisis de componentes principales". *Livestock Research for Rural Development*, 30(12), ISSN: 0121-3784, Available: <http://www.lrrd.org/lrrd30/12/arel30197.html>, [Consulted: January 3rd, 2019].
- Kaiser, H.F. 1960. "The application of electronic computers to factor analysis". *Educational and Psychological Measurement*, 20(1): 141-151, ISSN: 1552-3888, DOI: <https://doi.org/10.1177/001316446002000116>.
- Kennedy, B.W., Schaeffer, L.R. & Sorensen, D.A. 1988. "Genetic properties of animal models. Proceedings of the animal model workshop". *Journal of Dairy Science*, 71(2): 17-26, ISSN: 0022-0302, DOI: [https://doi.org/10.1016/S0022-0302\(88\)79975-0](https://doi.org/10.1016/S0022-0302(88)79975-0).
- Khan, T.A., Tomar, A.K.S., Dutt, T. & Bhushan, B. 2013. "Principal component regression analysis in lifetime milk yield prediction of crossbred cattle strain Vrindavani of North India". *Indian Journal of Animal Sciences*, 83(12): 1288-1291, ISSN: 2394-3327.
- Ojango, J.M., Ducrocq, V. & Pollott, G.E. 2005. "Survival analysis of factors affecting culling early in the productive life of Holstein-Friesian cattle in Kenya". *Livestock Production Science*, 92(3): 317-322, ISSN: 0301-6226, DOI: <https://doi.org/10.1016/j.livprodsci.2004.08.011>.
- Pesek, J. & Baker, R.J. 1969. "Desired improvement in relation to selection indices". *Canadian Journal of Plant Science*, 49(6): 803-804, ISSN: 1918-1833.
- Ponce de León, R. & Guzmán, G. 1993. "Relationship between survival of cows and milk production during first lactation. 1. Genetic correlations and correlations between breeding values". *Cuban Journal of Agricultural Science*, 27(1): 17-24, ISSN: 2079-3480.
- Ponce de León, R., Ribas, M., Guzmán, G., Gutiérrez, M. & Mora, M. 2002. Longevidad, supervivencia y producción por vida de la nueva raza Siboney de Cuba. *Memorias II Congreso Internacional de Mejoramiento Animal*. 18-22 de noviembre. Palacio de las Convenciones, La Habana, Cuba.
- Rencher, A.C. 2002. *Methods of multivariate analysis*. 2nd Ed. Ed. Wiley-Interscience. New York, USA, p. 695, ISBN: 0-471-41889-7.
- Rouvier, R. 1969. "Ponderation des valeurs genotypiques dans la selection par indice sur plusieurs caractères". *Biometrics*, 25(2): 295-307, ISSN: 0006-341X, DOI: <https://doi.org/10.2307/2528790>.
- Ruales, F.R., Manrique, C. & Cerón, M.F. 2007. *Fundamentos en mejoramiento animal*. 1st Ed. Ed. L. Vieco e Hijos Ltda., Medellín, Antioquia, Colombia, p. 146.
- Ruiz, R., Blake, R.W., Castro, H.M.A., Sánchez, F., Montaldo, H.H. & Castillo, H. 2007. "Short communication: Changes in the association between milk yield and age at first calving in Holstein cows with herd environment level for milk yield". *Journal of Dairy Science*, 90(10): 4830-4834, ISSN: 0022-0302, DOI: <https://doi.org/10.3168/jds.2007-0156>.
- SAS (Statistical Analysis Systems). 2013 *The SAS System for Windows*, Release 9.3. Statistical Analysis Systems Institute, Cary, North Carolina, USA.
- Smith, H.F. 1936. "A discriminant function for plant selection". *Annals of Eugenics*, 7(3): 240-250, ISSN: 1469-1809, DOI: <https://doi.org/10.1111/j.1469-1809.1936.tb02143.x>.
- SPSS. 2002. *Statistical Package for Social Sciences. User's Guide*. SPSS Inc. Chicago, Illinois, USA.
- Unalan, A. & Cebeci, Z. 2004. "Estimation of genetic parameters and correlations for the first three lactation milk yield in

Holstein Friesian cattle by the REML method". Turkish Journal of Veterinary and Animal Sciences, 28(6): 1043-1049, ISSN: 1303-6181.

Yamada, Y., Yokouchi, K. & Nishida, A. 1975. "Selection index when genetic gains of individual traits are of primary concern". The Japanese Journal of Genetics, 50(1): 33-41, ISSN: 1880-5787, DOI: <https://doi.org/10.1266/jjg.50.33>.

Received: April 15, 2020

Accepted: July 31, 2020