P-ISSN: 2304-3075; E-ISSN: 2305-4360
International Journal of Veterinary Science
www.ijvets.com; editor@ijvets.com

Research Article

Dimension Reduction of Phenotypic Yield and Fertility Traits of Holstein-Friesian Dairy Cattle using Principle Component Analysis

Sherif A Moawed[1] and Mohamed M Osman[2]

[1]Department of Animal Wealth Development, Biostatistics Division, Faculty of Veterinary Medicine, Suez Canal University, Ismailia, 41522, Egypt; [2]Department of Animal Wealth Development, Faculty of Veterinary Medicine, Suez Canal University, Ismailia, 41522, Egypt
***Corresponding author:** sherifstat@yahoo.com; sherifmoawed@vet.suez.edu.eg

Article History: Received: February 01, 2018 Revised: April 15, 2018 Accepted: May 05, 2018

ABSTRACT

This study was undertaken to explain the variability in milk production and fertility traits of Holstein-Friesian cows by a reduced dimension using principle components analysis (PCA). A total of 3513 lactation records were analyzed covering the period from 2009 to 2017. The traits measured were; milk yield (MY), days in milk (DIM), days dry (DD), open days (OD), age at calving (AAC), calving interval (CI), services per conception (SPC), days in milk to first heat (DIMFH) and days in milk to first breed (DIMFB). Two datasets were used in this investigation, handling a number of the first lactation and pooled lactation traits. The sampling adequacy measures were verified, where the Kaiser-Meyer-Olkin (KMO) estimates were above 0.7, also the Bartlett's test denoted significant ($P<0.01$) outcomes. Three principle components were retained and rotated, elucidating 73% of traits variation. The first principle component (PC1) loaded heavily for MY, DIM, OD and SPC. PC2 had high loadings with DD, AAC and CI, while the third PC correlated mostly with DIMFH and DIMFB. The efficacy of PCA was confirmed by high communality estimates and the superiority of PC scores over original traits through stepwise regression analyses. These findings suggest that selection indices and breeding schemes for the current herd could be structured effectively using only three components instead of nine original traits, without significant loss of information.

Key words: Milk yield, Fertility traits, Principle components analysis, Multivariate methodology, Holstein-Friesian cows

INTRODUCTION

Measuring phenotypic traits in dairy cattle herds is the substantial step for providing information about trait variations and efficiently planning selection programs. Breeding strategies that have been established on the basis of single-trait selection lead to misunderstanding of the actual herd performance with considering biological and genetic relationships between lactation and fertility traits (Rosario *et al*., 2008). Accordingly, the implementation of univariable statistical techniques had become disadvantageous and inappropriate to explain the maximal amount of variability in animal models. Alternatively, multivariate approaches have been recommended by previous literatures (Karacaoren and Kadarmideen 2008 and Angelina *et al*., 2016) as more beneficial tools for analyzing dairy records. However, these multivariate methodologies constitute a problem represented in the inherent correlations that could be noticed among the investigated traits. In other words, most of multivariate models require the explanatory traits not to be highly correlated to minimize the potential occurrence of multicolinearity. Because most of multivariate methods are multiple regression based models, the existence of collinearity among the independent variables could lead us to violation of assumptions such as linearity, normality, homoscedasticity and independence of studied variables (Karacaoren and Kadarmideen 2008). Data dimension, which reflect the number of traits incorporated in the analytical models, have to take in consideration when dealing with multivariate animal models. Because animal breeding strategies and selection indices have frequently been constructed with big datasets, researchers may not able to determine the unnecessary traits, or could have traits that can explain little variability.

Principle components analysis (PCA) is a data reduction multivariate approach that can be used to minimize a large number of variables into smaller sets. The main idea of this technique is to pick up the correlated variables together in the form of orthogonal or

Cite This Article as: MoawedSA and MM Osman, 2018. Dimension reduction of phenotypic yield and fertility traits of holstein-friesian dairy cattle using principle component analysis. Inter J Vet Sci, 7(2): 75-81. www.ijvets.com

independent clusters or groups, called principle components (PCs). These PCs are linearly combined with the original data, and characterized by its ability to retain all information denoted by the observed and latent variables (Egena*et al*., 2014). Therefore, this technique can overcome the collinearity problem by breaking the potential dependency among the observed variables. PCA has been approved by Taggar (2011) as a dimension reduction method, because it can be used to summarize the original data into uncorrelated principle components. The first few PCs elucidate the greatest amount of variation in the analyzed dataset. There have been a number of studies involving the use of PCA to analyze functional and economic traits of different dairy and beef cattle breeds. Most of the previous studies (Meyer, 2005; Savegnago*et al*., 2011; Bignardi*et al*., 2012; Boligon*et al*., 2013; Agudelo-Gomez *et al*., 2015) have been conducted to estimate genetic parameters for genetic improvement of dairy and beef cattle herds. In all previous investigations, PCA was proved to be beneficial for reduction of the original datasets, allowing reasonable estimates of genetic parameters. In term of accuracy, PCA was reported as a more effective methodology than the multiple regression models for early selection and prediction of milk production in Friesian breeds (Chakravarty *et al*., 1998).

In this study, principle components analysis was applied in an attempt to reduce the data dimension of milk yield and fertility traits of Holstein-Friesian cows using first lactation and pooled lactation traits. In addition, this work was planned to use the multiple regression analysis to evaluate the accuracy of PCA, by modeling milk yield with the original traits and with principle component scores.

MATERIALS AND METHODS

Data source, herd management and measured traits

Data of the present study were obtained from 3513 lactation records related to Holstein-Friesian dairy cows stationed at Damietta Governorate of Egypt, over the period of 2009-2017. All over the year, all animals were raised either on dirty floor system in open yards or in partially protected open yards along with cool spraying system through hot conditions. The feed for cows contained 18 – 19 % crude protein according to the National Research Council. Cows were milked three times daily with 8 hours interval, using automatic milking systems in herringbone parlor. Computerized recording of data was carried out using different types of electronic systems such as Afikim and Dairy Comb 305. Heifers were artificially inseminated using frozen semen collected from Holstein bulls in Canada and U.S.A. based on the total predicted index (TPI). The present study was carried out using dataset of the first lactation (n=1963), another analysis was achieved using pooled lactations dataset, representing the first six lactations (n=3513). A number of milk yield and fertility traits were considered in this study; lactation milk yield (LMY), days in milk (DIM), number of dry days (DD), open days (OD), age at calving (AAC), calving interval (CI), services per conception (SPC), days in milk to first heat (DIMFH) and days in milk to first bred (DIMFB). Further analyses were undertaken using first lactation and pooled lactation datasets in an attempt to compare the results, which could be useful for breeding purposes.

Principle components analysis

Prior to statistical analyses of the current datasets using principle components analysis (PCA), data have been carefully checked for the existence of outliers, missing values and have also been tested for the normality assumption of multivariate methodology. Interestingly, the eligibility criteria required by PCA have been examined, particularly, the sampling adequacy and the correlation matrix of investigated traits. Kaiser-Meyer-Olkin (KMO) sampling adequacy measure (Cerny and Kaiser, 1977) was computed to verify the relevance of the studied datasets to PCA. The KMO measured the magnitude of partial correlations among traits. The estimate of KMO ranged from 0 to 1 where value greater than or equal to 0.7 was considered good (Hutcheson and Sofroniou, 1999). In a similar manner, Eyduran*et al*. (2010) deemed the KMO above 0.6 was adequate for PCA. The KMO estimate of sampling adequacy was denoted by the formula:

$$KMO = \frac{\sum_{i \neq j} R_{ij}^{\ 2}}{\sum_{i \neq j} R_{ij}^{\ 2} + \sum_{i \neq j} C_{ij}^{\ 2}} \quad (1)$$

Where R_{ij} is the correlation matrix and C_{ij} is the partial covariance matrix.

The Bartlett's test of sphericity (Snedecor and Cochran, 1989), another overall measure of sampling adequacy, was applied to check the suitability of data to be reduced using PCA. This test compared the correlation matrix of measured traits with the matrix of zero correlations, which was known as the identity matrix to examine the overall relation among traits. In term of statistical inference, Bartlett's test of sphericity has been computed to test the null hypothesis that the correlation matrix wasn't diverge from the identity matrix (under Ho: traits are orthogonal). The PCA can perform a reduction of traits dimension without loss of information only if the null hypothesis was rejected ($p<0.05$). The Bartlett's test of sphericity, which follows a chi-square distribution with a [p (p-1) / 2] degree of freedom is given as:

$$\chi^2 = (1 + \frac{2p+5}{6} - n) \, Log|R| \quad (2)$$

Where p is the number of traits, n is the overall sample size, and |R| is the determinant of correlation matrix R. Moreover, the sampling sufficiency of every trait was identified using the partial correlations, namely, anti-image correlations, which were originated from the R matrix. The multicolinearity among studied traits was detected by calculating the determinant scores of the correlation matrix. As a general rule, determinant scores greater than 0.00001 imply the absence of traits collinearity (Haitovsky, 1969; Field *et al*., 2012).

Understanding the investigation, PCA is a multivariate method used to explain the total variability in datasets by transforming a set of correlated traits, X_1, X_2, …, X_p into new orthogonal and uncorrelated variables

called principle components, PC_1, PC_2, …, PC_p. The new extracted components, which are linear combination of the original observed traits, can be expressed in the following equations:

$$PC_1=a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$
$$PC_2=a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \quad (3)$$
$$PC_p=a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p$$

The general form of any extracted PC_i can be written as follow:

$$PC_i=a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{ip}X_p \quad (4)$$

Where a_{i1}, $a_{i2,}$ … $a_{ip}$are the component coefficients, or the loadings traits on each principle component, with i=1, 2, …, p. According to Everitt*et al.* (2001), the first PC denotes the highest percentage of explained variability in datasets, followed by PC_2 and then the other PCs. In this study, the extracted components have been rotated using varimax rotation with Kaiser Normalization. The eigenvalues of each component were estimated and arranged in descending order, revealing the proportions of variance in the original traits. The Communality (C_j^2) of each PC, which was estimated by summing the squared loadings of each component, is given as follows:

$$C_j^2=a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2 \quad (5)$$

In the end, stepwise multiple regression analysis was carried out using the original traits as the explanatory variables in standardized form, and another analysis based on the principle component scores. In both analyses, coefficients of determination were evaluated and compared. All statistical analyses were performed using SPSS for windows (SPSS, 2007) and the PRINCOMP procedure of SAS version 9.2. (SAS institute, 2008).

RESULTS AND DISCUSSION

The preliminary results obtained from principle components analysis are summarized in Table 1. The data included means, standard deviations and coefficient of variations of the studied traits for both the first lactation and pooled lactations. The phenotypic means of studied traits in early lactation were 7521.1, 251.2, 67.3, 174.9, 35.4, 462.2, 3.8, 50.3 and 71.3 for LMY, DIM, DD, OD, AAC, CI, SPC, DIMFH and DIMFB, respectively. The corresponding means of these traits for pooled lactations were similar to some extent to those of the first lactation. The relative standard deviations or coefficients of variation (ranged from 24.3 % to 85.3 %) revealed high variability for most of traits, both for animals in early stage of production and for accumulated lactations. The results in Table 1 indicate the importance of the studied traits in planning for dairy cattle breeding programs, starting from the early stage of milk yield. The correlation coefficients among milk yield traits and fertility traits are set out in Table 2. The data of the first lactation and pooled lactations revealed that LMY was positively correlated with all other traits, except for DD. The highest correlations (≥0.69) with LMY were recorded for DIM, OD and SPC. Out of the total existed 36 correlations for first lactation, there were 22 that were significant ($P\leq0.05$).

In term of the goodness of fit the principle components analysis to data of this study, Table 3 illustrates the measures of sampling adequacy, sufficiency proof for the validity of PCA to the current datasets, along with multicolinearity diagnostics. Interestingly, the Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy were 0.720 and 0.723, for the first lactation and pooled lactation traits, respectively. Comparatively, the current values of KMO were similar to the estimate (0.71) reported by Dhakal, (2017) and close to the value (0.75) of Verma*et al*. (2015), higher than earlier estimate (0.60) of Tolenkhomba*et al*. (2013) in hill cattle, whereas in other studies (Pundir*et al*., 2011; Egena*et al*., 2014), higher estimates of KMO (0.81) were recorded. The KMO estimates in this study indicated that PC methodology was appropriate for data reduction and the overall sample sizes were sufficient enough to get reliable estimates. The anti-image partial correlations on the diagonal of correlation matrix (Table 2) were all greater than 0.5, suggesting the sampling adequacies (Dhakal, 2017) for each individual trait, being analyzed using either the first lactation or pooled lactation datasets. Moreover, the results of Bartlett's test of sphericity (Table 3) were highly significant for early lactation traits (chi-square=9962.2, $P<0.001$) and for pooled lactation traits (chi-square=16888.65, $P<0.001$), providing more evidences for the validity of PCA. Accordingly, the results of Bartlett's test support the alternative hypothesis that the original correlation matrix is not an identity matrix, confirming the existence of relationships among traits. The results of determinant scores (0.01 and 0.008), as shown in Table 3, were greater than 0.00001, indicating the absence of traits multicolinearity (Field *et al*., 2012) for all analyzed datasets. Taken together, these results provide important proofs for the compatibility of PCA to the current dairy datasets.

Table 1: Mean, standard deviation (SD) and coefficient of variation (C.V. %) for the first lactation and pooled lactations traits.

Trait	First lactation (n = 1963)			Pooled lactations (n = 3513)		
	Mean	SD	C.V. %	Mean	SD	C.V. %
Lactation milk yield (LMY, kg)	7521.1	4491.4	59.7	7602.9	4357.9	57.3
Days in milk (DIM)	251.2	156.9	62.5	247.1	105.2	42.6
Dry days (DD)	67.3	17.1	25.4	67.4	36.7	54.5
Open days (OD)	174.9	135.4	77.4	172.4	94.6	54.8
Age at calving (AAC, month)	35.4	16.1	45.5	48.3	24.9	51.6
Calving interval (CI, days)	462.2	120.9	26.2	456.7	131.3	28.7
Number of services per conception (SPC)	3.8	3.24	85.3	3.7	3.1	83.8
Days in milk to first heat (DIMFH)	50.3	22.3	44.3	49.1	22.2	45.2
Days in milk to first breed (DIMFB)	71.3	17.6	24.7	70.5	17.1	24.3

Table 2: Correlation coefficient between the first lactation traits (above diagonal), between pooled lactation traits (below diagonal), and anti-image correlations (on diagonal).

Trait	LMY	DIM	DD	OD	AAC	CI	SPC	DIMFH	DIMFB
LMY	0.78(0.79)	0.90**	- 0.03	0.82**	0.06**	0.03	0.69**	0.07**	0.14**
DIM	0.88	0.76 (0.77)	- 0.03	0.84**	0.01	- 0.02	0.69**	0.09**	0.16**
DD	- 0.03*	- 0.03*	0.52 (0.49)	- 0.03	0.22**	0.70**	- 0.02	- 0.04*	- 0.08**
OD	0.79	0.83	- 0.02	0.79 (0.78)	0.02	0.03	0.80**	0.16**	0.21**
AAC	0.06	0.02	0.06	0.05	0.66 (0.47)	0.31**	0.03	- 0.12**	- 0.02
CI	0.07	0.03*	0.70	0.08	0.12	0.51 (0.52)	0.02	0.04*	- 0.02
SPC	0.69	0.69	- 0.02	0.81	0.04	0.06	0.80 (0.79)	0.05**	0.02
DIMFH	0.06	0.09	- 0.03*	0.15	- 0.08	0.06	0.03*	0.54 (0.52)	0.27**
DIMFB	0.13	0.15	- 0.04*	0.21	0.02	0.02	0.02	0.28	0.53 (0.51)

First lactation (n=1963) and pooled lactations (n = 3513); *Significant at 0.05 level (P<0.05); **Significant at 0.01 level (P<0.01).

Table 3: Measures of sampling adequacy and suitability indices for principle components analysis (PCA) for the first and pooled lactations traits.

Dataset	Measures of sampling adequacy	Statistics
First lactation traits	Kaiser-Meyer-Olkin measure of sampling adequacy	0.720
	Bartlett's test of sphericity:	
	Chi-square statistics	9962.2
	Df	36
	P value for significance	0.0001
	Determinant score	0.01
Pooled lactation traits	Kaiser-Meyer-Olkin measure of sampling adequacy	0.723
	Bartlett's test of sphericity:	
	Chi-square statistics	16888.65
	Df	36
	P value for significance	0.0001
	Determinant score	0.008

The eigenvalue, percentage of variance accounted for by each PC and the accumulated proportions of explained variances are presented in Table 4. This table shows the initial solution for PCA, which have been conducted on the first lactation and the pooled lactation datasets. Nine PCs have been denoted for both datasets, explaining the total percentage of variability in the original nine traits. The eigenvalues along with the corresponding percentages of variances were ranked in descending order, starting from 3.429 (38.1 %) for the first PC to 0.089 (0.987 %) for the last PC of the first lactation dataset. Similarly, the eigenvalues and proportions of variances explained for pooled lactation traits were 3.41 (37.89 %) for the first PC and 0.105 (1.168 %) for the last PC. Table 5 provides the results denoted by PCA extraction and varimax orthogonal rotation for the first lactation traits. Kaiser Rule criterion (Johnson and Wichern, 1998) retained only three PCs that have eigenvalues greater than one. In addition, the result of scree plot (Figure 1) revealed the total number of components along with the considered PCs for data reduction. In other words, scree plot has been applied to portray the components having eigenvalues up to the bent of elbow (>1.0) to be retained. It was apparent from Table 4 that the first three principle components explained about 73 % of the total variation in milk yield and fertility traits of the first lactation. The first three PCs explained about 38 %, 21 % and 14 %, respectively, of total variability in data of the early stage of productive life of cows. Strictly speaking, the percentage of variance (73 %) accounted for by extracted components regarded with the first lactation was acceptable. This is in agreement with Truxillo (2003) and Robin (2012) who stated that cumulative proportion of variance explained by most PCA was 70-80%.

As shown in Table 5, the first PC had high positive loadings on LMY (0.929), DIM (0.933), OD (0.932) and SPC (0.868). The second PC had high positive correlations with DD (0.872), AAC (0.532) and CI (0.905), while the third PC was highly loaded with positive correlations on DIMFH (0.807) and DIMFB (0.755). The communalities of first lactation traits (Table 5) were high and close to one, except for AAC (0.324). High communalities provide more credence to the effectiveness of PCA. According to Wuenseh (2012), communality is the coefficient of determination of the trait predicted from the PC. In similar studies (Yakubu *et al*., 2009; Ogah, 2011), high communalities have been estimated by the first PCs.

Regarding PCA of pooled lactation traits, the extracted components, loading of each PC and communalities are presented in Table 6. Three PCs have been extracted with eigenvalues greater than one (Figure 2), which jointly modeled about 71 % of total variation in the original dataset (Table 4). What was also important was that the three extracted components picked up the same traits as has occurred in the first lactation traits. The first component which modeled 37.88 % of variation had high positive loadings on LMY (0.921), DIM (0.928), OD (0.933) and SPC (0.873). PC2 which explained 19.17 % of lifetime variability loaded heavily on DD (0.910), AAC (0.221) and CI (0.921). The third PC was highly and positively correlated with DIMFH (0.806) and DIMFB (0.759). Moreover, the communalities of all traits were high, except for AAC. The results of communalities in this study are in accordance with Verma*et al*. (2015) who reported that communality estimates ranged from 0.41 to 0.88. The trend observed in the variability percentages explained by the rotated PCs came in agreement with most studies (Weigel and Rekaya, 2000; Kannan and Gandhi, 2004; Macciotta*et al*., 2006; Jaiswal *et al*., 2006; Haile *et al*., 2008; Angelina *et al*., 2016; Zefrehei*et al*., 2016). Data from Table 6 can be compared with the data in Table 5 which suggest that the early lactation traits could be used effectively in predicting the lifetime performance of dairy cows, both in milk yield and fertility measures. The unique factor presented in both Tables 5 and 6 denoted the unexplained variation revealed by each trait, which were all low, except for AAC. A similar study had been conducted by Angelina *et al.* (2016) who used PCA to reduce the datasets of milk production traits and reported that the original traits have been clustered into two components, explaining 89 % of the total variation in lactation traits.

Table 4: Initial eigenvalues, percentage of variance explained by each principle component (PC) along with the accumulated proportions of variances for the first and pooled lactations traits.

Principle Component	Using first lactation traits			Using pooled lactation traits		
	Eigenvalue	Percentage of variance	Accumulated percentages	Eigenvalue	Percentage of variance	Accumulated percentages
PC1	3.429	38.101	38.101	3.410	37.888	37.888
PC2	1.874	20.821	58.922	1.725	19.171	57.059
PC3	1.246	13.841	72.764	1.268	14.093	71.152
PC4	0.878	9.754	82.518	0.987	10.962	82.113
PC5	0.696	7.735	90.253	0.712	7.915	90.028
PC6	0.367	4.073	94.326	0.368	4.092	94.121
PC7	0.287	3.184	97.510	0.283	3.141	97.261
PC8	0.135	1.503	99.013	0.141	1.571	98.832
PC9	0.089	0.987	100.00	0.105	1.168	100.00

Table 5: Principle component loadings, communalities and uniqueness for the extracted and rotated PCs with the summary of the original traits (first lactation traits).

Traits	Extracted and rotated PCs[a]			Communality	Unique factor
	PC1	PC2	PC3		
LMY	0.929	0.023	0.055	0.867	0.133
DIM	0.933	- 0.013	0.088	0.879	0.121
DD	- 0.047	0.872	0.010	0.762	0.238
OD	0.932	0.016	0.161	0.895	0.105
AAC	0.063	0.532	- 0.192	0.324	0.676
CI	- 0.003	0.905	0.097	0.828	0.172
SPC	0.868	0.013	- 0.046	0.756	0.244
DIMFH	0.040	- 0.027	0.807	0.654	0.346
DIMFB	0.102	- 0.054	0.755	0.584	0.416

[a] Extraction Method: Principle Component Analysis; [a] Rotation Method: Varimax with Kaiser Normalization.

Table 6: Principle component loadings, communalities, uniqueness and the diagonal anti-image correlation for the extracted and rotated PCs with clustered traits (pooled lactations).

Traits	Extracted and rotated PCs[a]			Communality	Unique factor
	PC1	PC2	PC3		
MY	0.921	0.027	0.034	0.851	0.149
DIM	0.928	0.001	0.070	0.866	0.134
DP	- 0.058	0.910	-0.030	0.833	0.167
DO	0.933	0.039	0.138	0.892	0.102
AAC	0.078	0.221	- 0.207	0.098	0.902
CI	0.039	0.921	0.065	0.854	0.146
SC	0.873	0.028	- 0.069	0.768	0.232
DIMFH	0.044	0.016	0.806	0.652	0.348
DIMFB	0.155	- 0.007	0.759	0.590	0.41

[a]Extraction Method: Principle Component Analysis; [a]Rotation Method: Varimax with Kaiser Normalization.

Table 7: Stepwise regression for testing the relationship between milk yield and other traits using the original standardized traits of first lactation and pooled lactation traits.

Model	Model summary							
	First lactation				Pooled lactations			
	Predictor	R^2	SE	F	Predictor	R^2	SE	F
Using original traits (standardized)								
1	a	0.816	0.42	8678.4**	a	0.777	0.47	12255.8
2	b	0.827	0.41	4670.1**	b	0.792	0.45	6671.9
3	c	0.829	0.41	3173.5**	i	0.794	0.45	4513.2
4	d	0.831	0.41	2402.5**	d	0.796	0.45	3414.8
5	e	0.831	0.41	1928.1**	j	0.796	0.45	2741.3
6	f	0.832	0.41	1611.1**	k	0.797	0.45	2299.5
7	g	0.834	0.40	1402.5**	g	0.798	0.44	1980.2
8	h	0.834	0.40	1505.23	l	0.725	0.43	1785.1
Using principle component scores								
1	m	0.863	0.36	12385.9**	m	0.849	0.38	19729.3
2	n	0.866	0.36	6349.9**	n	0.850	0.38	9949.4
3	o	0.867	0.36	4250.1**	o	0.851	0.38	6668.9

Model contents of explanatory traits beside the constant term:a. (DIM), b. (DIM, DO), c. (DIM, DO, AAC), d. (DIM, DO, AAC, SC), e. (DIM, DO, AAC, SC, DIMFH), f. (DIM, DO, AAC, SC, DIMFH, CI), g. (DIM, DO, AAC, SC, DIMFH, CI, DP), h. (DIM, DO, AAC, SC, DIMFH, CI, DP, DIMFB), i. (DIM, DO, SC), j. (DIM, DO, SC, AAC, CI), k. (DIM, DO, SC, AAC, CI, DP), l. (DIM, DO, AAC, SC, DIMFH, CI, DP, DIMFB), m. (first PC scores), n. (first PC scores, third PC scores), o. (first PC scores, second PC scores, third PC scores).

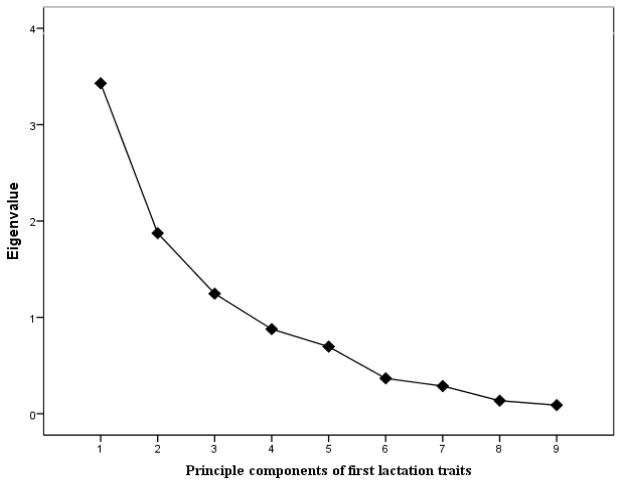

Fig. 1: Scree plot of principle components along with their eigenvalues for the first lactation traits.

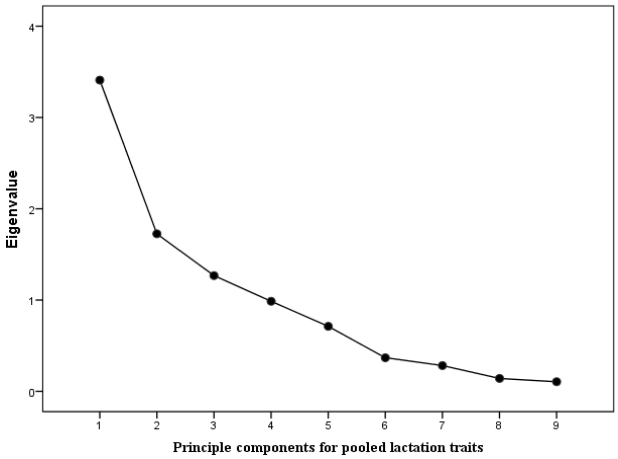

Fig. 2: Scree plot of principle components along with their eigenvalues for the pooled lactation traits.

The results of stepwise multivariable regression analyses (Table 7) denoted the model accuracy regarding predicting LMY from the measurements of the original traits and from principle component scores, using all studied datasets. Overall, the percentage of explained variation in milk yield predicted from the first lactation traits was 83 %, while the lifetime traits explained about 79 % of variability in LMY. On the other hand, an improvement was observed in the proportion of variance explained in models with PC scores. Modeling LMY with the first PC scores only accounted for 86.3 % of variability in MY in earlier stages of animal performance. Combination of component scores of PC1, PC2 and PC3 in the model explained much more variability (>85 %) in MY than that done by the original traits, for first and pooled lactations. This finding agrees with those reported by Vaidya (2002) and Egena*et al.* (2014) who revealed the effectiveness of PC scores in modeling the variability of body weight and milk yield as compared with using the original traits as predictors. The principle components are orthogonal, which implies the independency (zero correlations) between all PCs, hence, selection of animals on the basis of any PC does not affect the other. This might indicate the preference of PCA in constructing selection indices rather than using the original traits. This is because of the collinearity problem associated with incorporation of interdependent original measurements, which may lead to unstable and unreliable estimates of coefficients (Malau-Adul*iet al.*, 2004).

Conclusions

The present study was undertaken to examine the appropriateness of principle components analysis in reducing the number of milk yield and fertility traits being used for breeding purposes of dairy cattle. This study has found that generally three PCs have been extracted from the first lactation and pooled lactation datasets, explaining most of variability originated from many correlated traits without missing information. The high loading denoted by the same traits on each PC along with high communality estimates suggested the possibility of selection of animals based on clustered traits rather than isolated traits. The results of this study also indicated that the first lactation traits could be used as a good indicator for the lifetime performance of dairy cows. These findings provide an enhancement in understanding and evaluating the total variability recorded in many functional traits of dairy animals, allowing for amenable reduction in the number of traits taken into account in selection index and breeding programs of Holstein dairy cows.

Acknowledgments

We are grateful to thank Egyptian Knowledge Bank (EKB, www.ekb.eg) and Egyptian Specialized Presidential Council for Education and Scientific Research for English improvement of the manuscript.

REFERENCES

Agudelo-Gómez DA, R Pelicioni-Savegnago, ME Buzanskas, AS Ferraudo, D Prado-Munari and MF Cerón-Muñoz, 2015. Genetic principal components for reproductive and productive traits in dual-purpose buffaloes in Colombia. J Anim Sci, 8: 3801-3809.

Angelina F, S Darlim, S Fabiane, H Kuang, M Thomas and L Fernando, 2016. Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. Trop Anim Health Prod, 3: 533-538.

Bignardi AB, L El Faro, GJM Rosa, VL Cardoso, PF Machado and L Galvão de Albuquerque, 2012. Short communication: Principal components and factor analytic models for test-day milk yield in Brazilian Holstein cattle. J Dairy Sci, 95: 2157–2164.

Boligon AA, AB Bignardi, MEZ Mercadante, RB Lôbo and LG Albuquerque, 2013. Principal components and factor analytic models for birth to mature weights in Nellore cattle. Livestock Sci, 152: 135–142.

Cerny CA and HF Kaiser, 1977. A study of measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behavior Res, 12: 43-47.

Chakravarty AK, Gandhi, RS, Singh, A Gurnani, M 1998. Effectiveness of Principal Component Analysis as a criterion of selection. Annual Report, 99, National Dairy Res. Institute, Karnal.

Dhakal B, 2017.Using Factor Analysis for Residents' Attitudes towards Economic Impact of Tourism in Nepal. Int J Stat Appl, 7: 250-257.

Egena SSA, AT Ijaiya, DM Ogah and VE Aya, 2014. Principal component analysis of body measurements in a population of indigenous Nigerian chickens raised under extensive management system. Slovak J Anim Sci, 47: 77-82.

Everitt BS, S Landau and M Leese, 2001. Cluster Analysis. 4th ed., Arnold Publisher, London.

Eyduran E, M Topal and AY Sonmez, 2010. Use of factor scores in multiple regression analysis for estimation of body weight by several body measurements in brown trouts (Salmo truttafario). Int J Agric Biol, 12: 611- 615.

Field A, J Miles and Z Field, 2012. Discovering statistics using R. SAGE Publishing, London.

Haile A, BK Joshi, W Ayaleq, A Tegegne, A Singh and AK Chakravarty, 2008. Prediction of first lactation milk yield of Boran cattle and its crosses with Holstein Friesian in Central Ethiopia using multiple regression and principal components analysis. Indian J Anim Sci, 78: 66-69.

Haitovsky Y, 1969. Multicolinearity in regression analysis: A comment. Review of Economics and Statistics, 51: 486-489.

Hutcheson G and N Sofroniou, 1999. The multivariate social scientist: Introductory statistics using generalized linear models. London: Sage Publication.

Jaiswal UC, AS Khanna and ML Sangwan, 2006. Evaluation of sire on multiple traits through factor analysis. Indian J Anim Sci, 76: 337-339.

Johnson RA and DW Wichern, 1998. Applied Multivariate Statistical Analysis, 5th ed., Prentice Hall, Texas.

Kannan DS and RS Gandhi, 2004. Principal component analysis: A multivariate criteria of selection in Sahiwal cows. Indian J Anim Sci, 74:1160-1163.

Karacaoren B and HN Kadarmideen, 2008. Principal Component and Clustering Analysis of Functional Traits in Swiss Dairy Cattle. Turkish J Vet Anim Sci, 32: 163-171.

Macciota NPP, D Vicario and A Cappio-Borlino, 2006. Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. J Dairy Sci, 89: 3188-3194.

Malau-Aduli AEO, MA Aziz, T Kojina, T Niibayashi, K Oshima and M Komatsu, 2004. Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. J Anim Vet Adv, 3: 856-863.

Meyer K, 2005. Genetic principal components for live ultrasound scan traits of Angus cattle. Anim Sci, 81:337–345.

Ogah DM, 2011. Shared variability of body shape characters in adult Muscovy duck. Biotechnol Anim Husb, 27: 189-196.

Pundir RK, PK Singh, KP Singh and PS Dangi, 2011. Factor analysis of biometric traits of Kankrej cows to explain body confirmation. Asian-Australasian J Anim Sci, 24: 449-56.

Robin B, 2012. An introduction to principal component analysis and factor analysis using SPSS 19 and R, United Kingdom, pp: 1-24.

Rosario MF, MA Silva, AA Coelho, VJ Savino and CT Dias, 2008. Canonical discriminant analysis applied to broiler chicken performance. Anim, 2: 419 - 424.

SAS Institute. 2008. SAS/STAT® 9.2 User's Guide. SAS Inst. Inc., Cary, NC

Savegnago RP, SL Caetano, SB Ramos, GB Nascimento, GS Schmidt, MC Ledurand DP Munari, 2011. Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. Poultr Sci, 90: 2174–2188.

Snedecor GW and WG Cochran, 1989. Statistical method. 8th edition, Iowa State University Press.

SPSS. 2007. Statistical package for the social sciences. SPSS Inc., 444 Michigan Avenue, Chicago, IL60611, USA.

Taggar RK, 2011. Development of selection criteria for jersey and jersey crossbred cows using principal component analysis. Ph.D. thesis, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Chatha, Jammu, pp: 70-115.

Tolenkhomba TC, SN Singh and DC Konsam, 2013. Principal component analysis of body measurements of bulls of local cattle of Manipur, India. Indian J Anim Sci, 83: 281-284.

Truxillo C, 2003. Multivariate Statistical Methods: Practical Research Applications Course Notes. Cary, NC: SAS Institute.

Vaidya M, 2002. Genetic evaluation of crossbred cows in Kounkan region of Maharashtra state. PhD Thesis, National Dairy Research Institute, Deemed University, Karnal, Haryana, India.

Verma D, V Sankhyan, S Katoch and YP Thakur, 2015. Principal component analysis of biometric traits to reveal body confirmation in local hill cattle of Himalayan state of Himachal Pradesh, India. Vet World, 8: 1453-1457.

Weigel KA and R Rekaya, 2000. A Multiple-Trait herd cluster model for International dairy sire evaluation. J Dairy Sci, 83: 815-821.

Wuenseh KL, 2012. Principal component analysis-SPSS, pp: 1-15.

Yakubu AD Kuje and M Okpeku, 2009. Principal components as measure of size and shape in Nigerian indigenous chickens. Thai J Agric Sci, 42: 167-176.

Zefrehei GM, MR Behzadi, M Fayaz and S Sharifi, 2016. Association between calving interval and productive traits in dairy cattle over different inseminations using artificial neural network. Small Rumin Res, 3: 169-187.