rccp Revista Colombiana de Ciencias Pecuarias

Multivariate analysis of milk yield, lactation length, and calving interval in female buffaloes¤

Análisis multivariado de la producción de leche, duración de la lactancia y el intervalo entre partos en búfalas

Análise multivariada da produção de leite, duração da lactação e intervalo de parto em búfalas

Jailton S Bezerra Junior[1], Zoo, MSc; Fabiane L Silva[2], Zoo, PhD; Kuang Hongyu[3], Mat, PhD; Alberto G Couto[4], Eng Civ, MSc; Rosália B N Medeiros[1], Zoo, PhD; Filipe C T Araújo[1], Zoo, MSc; Luciano G Lima[1], Zoo, MSc; Angelina B Fraga[1*], Zoo, PhD.

[1]*Centro de Ciências Agrárias, Universidade Federal de Alagoas, Rio Largo, Brasil.*

[2]*Departamento Ciências Agrárias, Ambientais e Biológicas, Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Brasil.*

[3]*Departamento de Estatística, Universidade Federal de Mato Grosso, Cuiabá 78.060-900, Brasil.*

[4]*Búfalo Bill, São Luiz do Quitunde, Alagoas, Brasil.*

(Received: August 19, 2016; accepted: September 18, 2017)

doi: 10.17533/udea.rccp.v31n2a03

Abstract

Background: The global market has an increasing demand for buffalo by-products due to their unique nutritional characteristics. Many buffalo herds lack suitable breeding control programs, hindering the implementation of selection programs. **Objective:** To evaluate milk production per lactation (MP), lactation length (LL), and calving interval (CI) of a herd of crossbred Murrah buffaloes to support buffalo clustering according to their potential. **Methods:** Data from 543 lactations between 2002 and 2014 from 105 crossbred Murrah female buffaloes were used. Data were subjected to principal component analysis (PCA) and cluster analysis. **Results:** The first components (PCs) were responsible for 92.32% of the total variation, of which 61.45 and 30.87% were explained by the first (PC_1) and second (PC_2) components, respectively. The cluster analysis allowed three female buffalo groups according to their potentials. **Conclusion:** Buffalo farmers can make decisions on nutritional, reproductive management and cow culling based on grouping.

Keywords: *Bubalus bubalis, cluster analysis, crossbred Murrah buffalo, dendrogram, principal component.*

¤ To cite this article: Bezerra JS J, Silva FL, Hongyu K, Couto AG, Medeiros R BN, Araújo F CT, Lima LG, Fraga AB. Multivariate analysis of milk yield, lactation length, and calving interval in female buffaloes. Rev Colomb Cienc Pecu 2018; 31(2):103-109.

* Corresponding author: Angelina B Fraga. Centro de Ciências Agrárias, Universidade Federal de Alagoas, Rio Largo, Brasil. 57.100-100. Tel: +55-82-99341-3092. E-mail: angelina.fraga@gmail.com

Resumen

Antecedentes: La demanda de productos de leche de búfala en el mercado mundial viene creciendo en virtud de sus características nutricionales únicas. La mayoría de los hatos bufalinos no tienen adecuados controles, lo que dificulta la implementación de programas de selección. **Objetivo:** Analizar la producción de leche por lactancia (MP), duración de la lactancia (LL) y el intervalo entre partos (CI) para apoyar la clasificación de búfalos según su potencial. **Métodos:** Se recogieron 543 registros de lactancia de 105 búfalas Murrah mestizas entre los años 2002 y 2014. Los datos fueron sometidos a análisis de componentes principales (PCA) y análisis de conglomerados. **Resultados:** Los primeros componentes (PCs) fueron responsables del 92.32% de la variación total, de los cuales 61.45 y 30.87% fueron explicados por el primer (PC_1) y segundo (PC_2) componentes, respectivamente. El análisis de conglomerados permitió la formación de tres grupos de búfalas, según su potencial. **Conclusión:** Los productores pueden tomar decisiones específicas con respecto a la gestión nutricional, reproductiva, y el descarte de hembras bufalinas basados en estas agrupaciones.

Palabras clave: *análisis de conglomerado, Bubalus bubalis, búfalas Murrah mestizas, componentes principales, dendrograma.*

Resumo

Antecedentes: A demanda por produtos de origem do leite de búfalas no mercado mundial tem crescido em virtude de suas características nutricionais singulares. Grande parte dos rebanhos de búfalos não possuem controle zootécnicos adequados, dificultando a implementação de programas de seleção. **Objetivo:** Analisar conjuntamente as características produção de leite por lactação (MP), duração da lactação (LL) e intervalo de partos (CI) para subsidiar a classificação das búfalas de acordo com seu potencial. **Métodos:** Para esse estudo, 543 dados de lactação de 105 búfalas mestiças Murrah foram coletadas entre 2002 e 2014. Os dados foram analisados por análises de componentes principais (PCA) e análises de cluster. **Resultados:** Os primeiros componentes (PCs) foram responsáveis por 92.32% da variação total, dos quais 61.45 e 30.87% foram explicados pelo primeiro (PC_1) e segundo (PC_2) componentes, respectivamente. A análise de cluster permitiu a formação de três grupos de búfalas de acordo com o potencial das características estudadas. **Conclusão:** Baseado nesses agrupamentos, os produtores podem tomar decisões específicas quanto aos manejos nutricional e reprodutivo, e sobre o descarte de búfalas, de acordo com os agrupamentos.

Palavras-chave: *análise de cluster, Bubalus bubalis, búfalas mestiças Murrah, componente principal, dendrograma.*

Introduction

The global population of Buffaloes (*Bubalus bubalis*) reaches approximately 195 million, with India as the country with the largest number (110 million; Food and Agriculture Organization —FAO, 2014). Buffalo milk production has shown a significant increase since the emergence of industrial units dedicated to production of milk derivatives. One of the main reasons for this remarkable growth has probably been the unique quality and flavor of these products, particularly, mozzarella cheese. In addition, buffalo milk has greater content of major constituents and more nutritional advantages compared to bovine milk (Aspilcueta *et al*., 2010).

The economic return of buffalos depends on milk production and reproductive efficiency of the animals, and both are affected by the calving interval (Ramos *et al*., 2006). Although there is a collective effort towards registering all the productive and reproductive information, a large portion of the world's buffalo farming does not have suitable data control for estimating the accurate breeding value of the buffaloes. The lack of appropriate data banks hinders the implementation of selection programs, which are essential for achieving genetic improvement. This study proposes a method to assist the ranking of buffaloes, despite having a compromised pedigree data structure.

In general, data sets from the animals are evaluated using univariate analysis. This type of analysis is limited since it does not consider possible correlations between traits in the study (Fraga *et al*., 2015). Multivariate analysis, including principal components and cluster, is an efficient method for sorting buffaloes according to productive and reproductive potential to assist in the selection and culling of animals, and to allow adjusting the management. It is important to note that these techniques have been used in livestock production and genetic improvement to decrease the sample space of the traits, and to assemble homogeneous groups as well as for other purposes (Kadegowda *et al.,* 2008; Neser *et al*., 2008; González *et al*., 2011; Ventura *et al*., 2012; Vohra *et al*., 2015).

This study proposes a method to assist the ranking of buffaloes despite a compromised pedigree data structure. Therefore, the objective of this study was to jointly evaluate milk yield per lactation (MP), lactation length (LL), and calving interval (CI) in a herd of crossbred Murrah buffaloes to support the clustering according to their potential.

Materials and methods

Data and management of buffaloes

The survey was conducted with milk production information from a herd of crossbred Murrah buffaloes from Castanha Grande Farm, located in the city of São Luiz do Quitunde, in the northern coast of Alagoas state, Brazil, with geographic coordinates: Latitude 9° 10’ 06” and longitude 35° 33’ 40”. Milk yield from 109 female buffaloes was analyzed. The data bank included 543 milk production records obtained between 2000 and 2014. The number of lactations per female buffaloes varied between 2 and 14. The following characteristics were evaluated: Milk production/lactation (MP) in litters, lactation length (LL), and calving interval (CI) in days for crossbred Murrah buffaloes. Buffaloes were milked twice a day (in the morning and afternoon) with a mechanical bucket milker using a piped vacuum line, and restrained in tandem (capacity: 24 individual buffaloes). Buffaloes were bathed before milking, which is important to stimulate intestinal evacuation. This procedure contributes to keeping the milking room clean. In addition to contributing to milk quality, this practice reduces stress. Buffaloes were bred during lactation in a semi-intensive system. The animals were kept in *Braquiária humidícula* pastures and received urea-enriched silage of sugarcane plants (*Saccharum officinarum*), including the aerial part. In addition, mineral supplementation and balanced concentrate were provided.

Feed was offered in the following proportions: Buffaloes yielding between 5 to 7.5 L of milk/d received 1 Kg of ration/d; buffaloes yielding between 8 to 10.5 L of milk/d received 2 Kg/d of ration.

Statistical analysis

Principal component analysis. Standardized phenotypic values of all traits from the 105 buffaloes were used in the PCA. Phenotypic values were standardized with the following formula:

$$\frac{x-\overline{x}}{s}$$

Where:

z: Is the standardized value of.

$\overline{x}$: Is the average of a trait.

s: Is the standard deviation of the trait.

Principal Components (PC) are calculated through linear combinations of the original variables with eigenvectors. The absolute value of an eigenvector determines the importance of the traits in a principal component. Each eigenvector is calculated from an eigenvalue of the correlation matrix of the data ($p<0.001$), where eigenvalues are related to the variance of each principal component (Rencher, 2002). The first principal component (PC1) explains the highest percentage of the total variance, the second principal component (PC2) explains the second highest, and so on, until all variances are explained. In a data bank with p variables, the random vector:

$$x' = [x_1, x_2, \ldots, x_p]$$

Having correlation matrix R with eigenvalue-eigenvector pairs (), for i = 1,2,...,p, Rencher (2002).

Where:

$$\lambda_1 \geq \lambda_2 \geq \ \ldots \geq \ \lambda_p \geq 0$$

The i^{th} principal component is given by:

$$PC_i = \boldsymbol{e}_i^t x = e_{i1}x_1 + e_{i2}x + \ldots + e_{ip}x_p$$

Where:

e_{ip}: Is the p^{th} eigenvector;

x_p: Is the p^{th} value of the original variable.

The criterion of Jolliffe (1972, 1973) was used to select the main components that explain most of the variation in the data bank. This criterion discards PC with eigenvalues less than 0.7 because it only explains a small part of the total data variability.

Previous group division by hierarchical clustering

Hierarchical cluster analysis using the Ward's criterion was performed by finding a suitable number of groups within the population. The Ward's method consists of aggregating two clusters such that growth of within-inertia is minimized (minimizing the reduction of the between-inertia) at each step of the algorithm. Within-inertia characterizes the homogeneity of a cluster. The hierarchy is represented by a dendrogram which is indexed by the gain of within-inertia. As previously mentioned, the hierarchical clustering is performed on the principal components (Lê *et al*., 2008).

Choice of number of clusters from a dendrogram

The choice of the number of clusters was performed using the "FactorMineR" package (Husson *et al*., 2016). This choice is a fundamental issue and several approaches have been proposed. The number of groups can then be chosen according to the overall look of the dendrogram, the bar graph, the variance within groups and other criterions. These rules are often based on, implicitly or not, the growth of inertia. They suggest a division into Q groups when the increase of variance between (Q-1) and Q cluster is larger than Q and (Q+1) cluster. All analyses were implemented with computational algorithms available in the *R* software 3.0.1 (R core team, 2016).

Results

The average (± standard deviation —SD) of milk yield per lactation, lactation length, and calving interval were 2,197.30 Kg (± 410.96), 277.01 d (± 37.21) and 438.35 d (± 66.41), respectively. The MP/LL, MP/CI, and LL/CI correlations (p value) were 0.72 ($p<0.00$), 0.10 ($p<0.32$), and 0.35 ($p<0.00$), respectively. Based on these results obtained for the main components, their respective eigenvalues and percentages of the variance explained by each, the first two components were responsible for more than 92.32% total data variation (Table 1). The first PC was responsible for approximately 61.45% and the second was responsible for 30.87% of the total variation.

Table 1. Principal components and percentage of variance explained by the components of milk yield, lactation length, and calving interval of Murrah breeding Buffaloes.

Principal components	λ_i	PCV (%)	PCV (% accumulated)
PC_1	1.84	61.45	61.45
PC_2	0.93	30.87	92.32
PC_3	0.23	7.68	100.00

PC_1: First principal component; PC_2: Second principal component; PC_3: Third principal component; λ_i: Eigenvalues; PCV: Percentage of variance explained by the components.

The 1st and 2nd equations were assembled using the first two principal components: (1st equation; 2nd equation). The main component in the 1st equation is the result of the linear combination of all the variables studied with the weighting coefficients, MP and LL, those which presented highest coefficients. Therefore, the first component is called productive component. The 2nd equation shows that the CI is the most important trait for the second main component, which is called the reproductive component (Table 2).

Table 2. Weighting coefficients for milk production, lactation length and calving interval of Murrah breeding Buffaloes and correlation with the first two principal components.

Traits	Correlations		Weighting coefficients	
	PC_1	PC_2	PC_1	PC_2
MP	0.86	-0.41	0.63	-0.43
LL	0.94	-0.07	0.69	-0.08
CI	0.48	0.87	0.36	0.90

MP: Milk production; LL: Lactation length; CI: Calving interval; PC_1: First principal component; PC_2: Second principal component.

Figure 1 shows the major explanation of the total variation of the data by two main components represented in a cartesian plane. Figure 2 presents information from two multivariate analysis techniques: Main components and cluster analysis. The dendrogram (Figure 3) was obtained by hierarchical cluster analysis with the grouping of the Ward method based on the Euclidean distance, separating the buffaloes into 3 groups according to the studied traits.

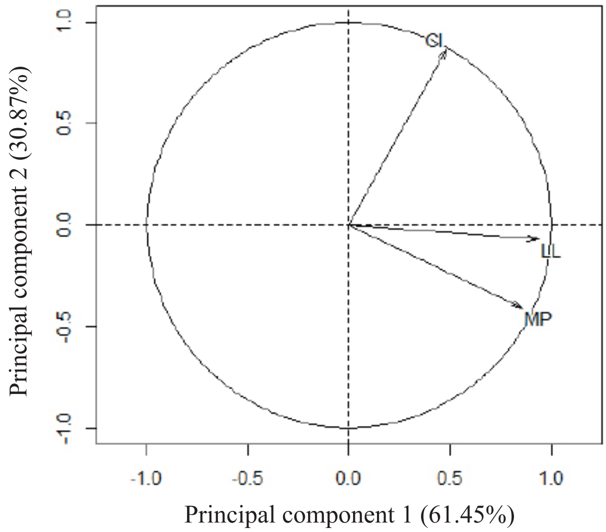

Figure 1. Incidence of vectors milk production (MP), lactation length (LL), and calving interval (CI) on the first two principal components.

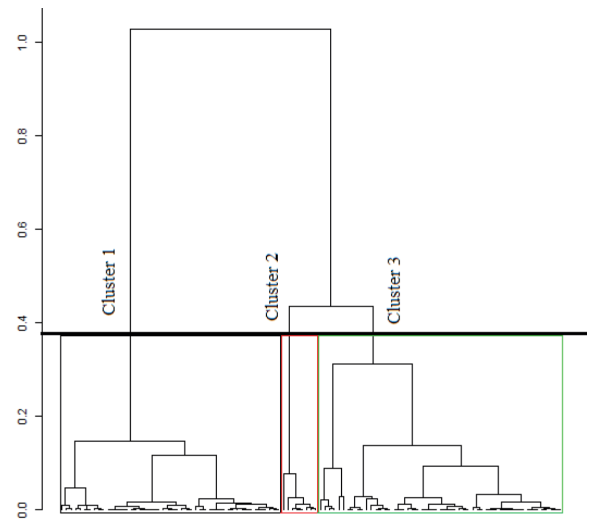

Figure 3. Buffaloes assembled by hierarchical clustering analysis according to MP (milk production per lactation), LL (lactation length), and CI (calving interval) performances.

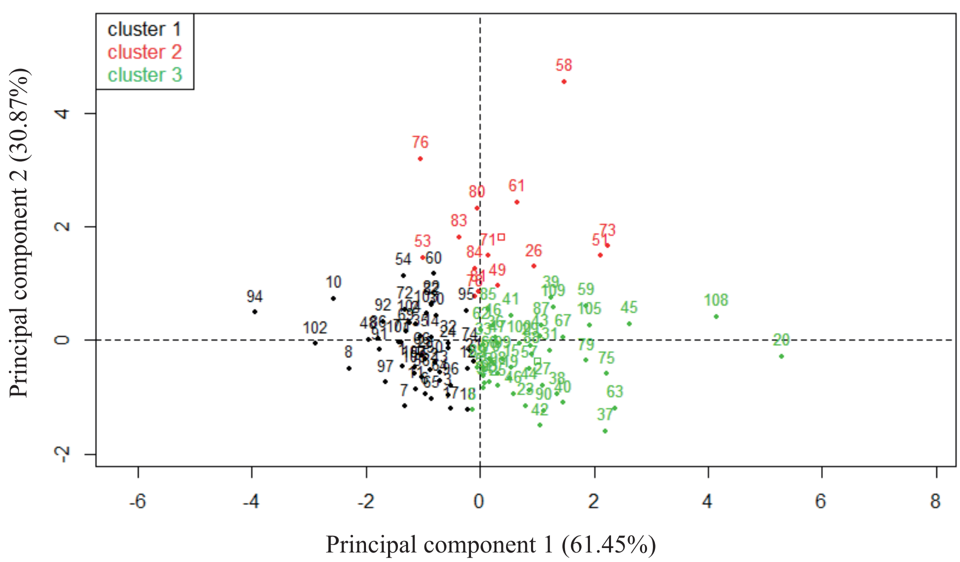

Figure 2. Grouping of female buffaloes in cluster 1, 2, and 3 according to the first two principal components.*

The average values for each one of these groups are shown in Table 3.

Discussion

Taking into account all factors in the production system, the herd showed superiority in milk production per lactation (2,232.86 Kg) and duration of lactation (275.44 d), in addition to lower range of calving interval (420.42 d), when compared to other herds. Rodrigues *et al.* (2010) studied 1,182 records of female Murrah Buffaloes and their cross-breedings in northern Brazil, finding average milk production per lactation and lactation length of 1,663.84 Kg and 269.89 d, respectively. Aspilcueta *et al.* (2010) estimated average 1,814 Kg for MP per lactation in southeastern Brazil. In northern Brazil, Pereira *et al.* (2008) found average of 451 d analyzing the information of 420 calving intervals in female Murrah x Mediterranean Buffaloes.

Table 3. Summary statistics for milk production (MP) per lactation (litters), lactation length (LL), and calving interval (CI) in days of breeding Murrah Buffaloes grouped in three clusters (Figure 2).

Cluster	Black (N = 47)		Red (N = 14)		Green (N = 95)	
variable	Average	(± SD)	Average	(± SD)	Average	(± SD)
MP	1,909.59	246.52	2,036.39	304.24	2,525.96	319.69
LL	251.01	25.35	275.53	30.27	302.89	30.64
CI	400.31	34.26	559.39	70.08	440.28	42.56

Correlation between milk production and lactation length (0.72) was high and positive, indicating that large LL will increases MP. Lower values for this association were reported by Malhado *et al*. (2009), Afzal *et al*. (2007) and Barros *et al*. (2016), who found values of 0.57, 0.62, and 0.59 in Murrah, Nili Ravi, and undefined breed buffaloes, respectively.

Correlation between MP and CI was low and positive (0.10). Therefore, the increase in MP leads to greater CI value. However, higher CI reflects low reproductive performance of buffaloes. Authors such as Ramos *et al*. (2006), and Malhado *et al*. (2009) found almost null correlation values between MP and CI, 0.01 and -0.02, respectively, indicating that these traits are phenotypically independent. That is, the variation of one does not imply altering the other.

The association between LL and CI was positive and median (0.35), showing that higher LL leads to increased CI, which is not desirable. In contrast, an increase in LL will contribute to an increase in MP, which is a favorable relation. Aziz *et al*. (2001) also found a positive and median correlation (0.47) between LL and CI in Egyptian buffaloes.

Considering the results of association between the traits studied, the use of principal components analysis can help interpreting the data, aiding in the decisions aimed to improve animal production. The PC technique allows creating a new set of uncorrelated variables from a set of correlated ones. This new set is able to explain the total variance observed between the data to facilitate the interpretation of interdependence between them.

The first two principal components effectively summarized total sample variance and it was possible to use the data bank study. In accordance with Jolliffe (1972; 1973), only PC3 component presented eigenvalue less than 0.7, which was disposed since it presented low importance to the data variability. Oliveira *et al*. (2014) applied the PCA technique to the genetic evaluation of 9 traits in Murrah Buffaloes finding that only the first 4 PCs were necessary to explain the covariance structure of the features. Furthermore, Agudelo-Gómez *et al*. (2015) reported that the first three PCs from eight reproductive features in dual purpose buffaloes in Colombia presented eigenvalues higher than one, which explained 65.8% of the total variation. Vohra *et al*. (2015) concluded that the first four components from 13 different body measurements of Gojri Buffaloes were sufficient to explain 70.9% of the total variation.

According to Rencher (2002), at least 70% of the total variation must be explained by the first two principal components. Hence, the first two principal components were enough for the quality of the hierarchical cluster analysis, through the dispersion of scores in graphics which axes are the components (Figure 2) under the conditions of our study.

Figure 2 shows the highest concentration of buffaloes of the green group in the same region where the LL and MP vectors (Figure 1), which exhibited high incidence on the first principal component. These buffaloes are those with the highest potential for LL and MP (Table 3). Regarding CI, the red group buffaloes are mostly found in the same quadrant as the CI vector (Figure 2), which has a higher impact on the second main component (Figure 1). Although the red group buffaloes presented high values for the CI, they do not represent the best performance for this trait. In this particular case, smaller values for the calving interval reflect the best results in a production system.

The green group gathered the buffaloes with the highest milk production per lactation and lactation length. In the red group, the buffaloes have the highest CI, and in this case, the worst performance for CI. On the other hand, the black group gathered buffaloes that showed inferior values of MP and LL, but intermediate values of CI, as seen in Table 3. These results provide buffalo classification that simultaneously takes into account the most important features of the system. The cluster analysis allowed three distinct buffalo groups according to milk production per lactation (Kg), lactation length (days), and calving interval (days), jointly evaluated. Based on buffalo grouping, farmers can take specific decisions on nutritional and reproductive management and culling. Therefore, when animals are separated in groups according to their productive and reproductive potential, they are also separated according to their nutritional or sanitary requirements. This technique also helps with reproductive management during the formation of the mating groups according to farm objectives.

Acknowledgments

The authors would like to thank Castanha Grande farm for providing the data for this study and CAPES/ Brazil (Coordenação de Aperfeiçoamento Pessoal de Nível Superior/Brazil) for financial support.

Conflicts of interest

The authors declare they have no conflicts of interest with regard to the work presented in this report.

References

Agudelo-Gómez D, Pineda-Sierra S, Cerón-Muñoz MF. Genetic Evaluation of dual-purpose buffaloes (*Bubalus bubalis*) in Colombia using principal component analysis. PloS ONE 2015; 10:1-9.

Afzal M, Anwar M, Mirza MA. Some factors affecting milk yield and lactation length in Nili Ravi buffaloes. Pakistan Vet J 2007:113-117.

Aziz MA, Schoeman SJ, Jordaan GF, El-Chafie OM, Mahdy AT. Genetic and phenotypic variation of some reproductive traits in Egyptian buffalo. SASAS 2001; 31(3):195-199.

Barros, CC, Aspilcueta-Borquis, RR, Fraga, AB, Tonhati, H. Genetic parameter estimates for production and reproduction traits in dairy buffaloes. Rev Caatinga 2016; 29(1):216-221.

Borquis RRA, Araújo Neto FR, Baldi F, Bignardi AB, Albuquerque LG, Tonhati H. Genetic parameters for buffalo milk yield and milk quality traits using Bayesian inference. J Dairy Sci 2010; 93:2195-2201.

Food and Agriculture Organization of the United Nations – FAO. Live Animals 2014.

Fraga AB, Silva FL, Hongyu K, Santos DDS, Murphy TW, Lopes FB. Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. Trop Anim Health Prod 2016; 48:533-538.

González A, Luque M, Rodero E, González C, Aguilera R, Jiménez J, Sepúlveda N, Bravo S, Herrera M. Use of morphometric variables for differentiating Spanish hound breeds. Int J Morphol 2011; 29:1248-1255.

Husson F, Josse J, Lê S. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining/R package version 1.33 2016

Jolliffe IT. Discarding Variables in a Principal Component Analysis. I: Artificial Data. J Roy Stat Soc 1972; 21:160-173.

Kadegowda AKG, Piperova LS, Erdman RA. Principal component and multivariate analysis of milk long-chain fatty acid composition during diet-induced milk fat depression. J Dairy Sci 2008; 91:749-759.

Jolliffe IT. Discarding Variables in a Principal Component Analysis. II: Real Data. J Roy Stat Soc 1973; 22:21-31.

Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw 2008; 25:1-18.

Malhado CHM, Ramos AA, Carneiro PLS, Azevedo DMMR, Affonso PRAM, Pereira DG, Souza JC. Estimativas de parâmetros genéticos para características reprodutivas e produtivas de búfalas mestiças no Brasil. Rev Bras Saúde Prod An 2009; 10(4):830-839.

Neser FWC, Erasmus GJ, Scholtz MM. The use of a cluster analysis in across herd genetic evaluation for beef cattle. South African J Anim Sci 2008; 38:51-57.

Oliveira DP, Barros CC, Araújo Neto FR, Lourenco DAL, Hurtado-Lugo NA, Tonhati H. Principal Components for Reproductive and Productive Traits in Buffaloes from Brazil. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Aug 17-22, Vancouver, Canada; 2014.

Pereira RGA, Barbosa SBP, Lopes CRA, Santoro KR, Townsend CR, Magalhães JA, Silva Neto FG, Costa NL. Intervalo de partos em rebanho bubalino no Estado de Rondônia. 53th Boletim de Pesquisa e Desenvolvimento, Embrapa Rondônia, Porto Velho, RO, Brazil; 2008.

RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, USA; 2016.

Ramos AA, Malhado, CHM, Carneiro, PLS, Gonçalves, HC, Azevedo, DMMR. Caracterização fenotípica e genética da produção de leite e do intervalo entre partos em bubalinos da raça Murrah. Pesquisa Agropecuária Brasileira 2006; 41(8):1261-1267.

Rencher A. Methods of multivariate analysis. 2nd ed. John Wiley & Sons, Inc, New York, NY, USA; 2002.

Rodrigues AE, Marques JRF, Araújo CV, Camargo Jr RNC, Dias LNS. Estimação de parâmetros genéticos para características produtivas em búfalos na Amazônia Oriental. Arq Bras Med Vet Zoo 2010; 62:712-717.

Ventura HT, Lopes PS, Peloso JV, Guimarães SEF, Carneiro APS, Carneiro PLS. Use of multivariate analysis to evaluate genetic groups of pigs for dry-cured ham production. Livest Sci 2012; 148:214-220.

Vohra V, Niranjan SK, Mishra AK, Jamuna V, Chopra A, Sharma N, Jeong DK. Phenotypic characterization and multivariate analysis to explain body conformation in lesser known buffalo (*Bubalus bubalis*) from North India. Asian-Aust J Anim Sci 2015; 28:311-317.