PCA-driven approach for prediction of body weight in Palla strain of Nellore sheep

K.E. Bhavani[1]*, S. Vani[1], K.S. Devi[1], J.M. Bhumireddy[2], P.P.R. Reddy[1], S.M. Siraj[1]

[1]Department of Animal Genetics and Breeding, College of Veterinary Science, Proddatur-516360, Sri Venkateswara Veterinary University, Andhra Pradesh; [2] Department of Animal Genetics and Breeding, School of Veterinary and Animal sciences, Centurion University of Technology and Management, R. Sitapur, Paralakhemundi, Gajapati, Odisha-761211 India.
*Corresponding author's E-mail: eswarkillana97@gmail.com

Journal of Livestock Science (ISSN online 2277-6214) 17: 181-188
Received on 4/5/25; Accepted on 26/3/26; Published on 30/3/26
doi. 10.33259/JLivestSci.2026.181-188

Abstract

Present investigation was carried to assess the morphometric structure of Nellore Palla sheep and to validate the relationship between body measures (BM) and body weight, utilising original BM and principal component scores. Principal component analysis (PCA) a multivariate statistical technique that transforms correlated variables into a smaller number of uncorrelated variables called principal components, was used to reduce dimensionality and improve prediction accuracy. The data comprising of body weight and eight body measurements from a total of 560 animals at age group of two teeth and above collected from the farmer flocks in Nellore district of Andhra Pradesh were utilised in the present study. The phenotypic correlations (r) for all BM with BW were positive and significant which ranged from 0.227 (Face width) to 0.938 (Chest girth). The PCA of BM and BW produced two components that accounted for 76.55% of the total variance. The first factor had higher values for variables related to body size, whereas second factor was loaded in favour of body shape. The prediction model for body weight was found to best with two principal components which produced higher R^2 value and lower RMSE, AIC, and BIC values, and when compared to models with one PC or based on original measurements. Our findings suggested that PC approach is a viable alternative for predicting BW and could be useful in designing management, selection and breeding programmes of Palla sheep.

Keywords: Body measurement; Palla Nellore sheep; Morphometric structure; Multi Variate analysis

Introduction

The morphological structure of livestock is accurately delineated by a combination of body measurements and body weight, which together provide a complete profile of species or breed. These linear body measurements are utilized in predicting body weight at field conditions (Mavule et al., 2013). Many researchers used classical principal component analysis (PCA) to extract factors contributing towards variation amongst individual animals based on body measurements (Mishra et al., 2016; Yadav et al., 2016,) in sheep breeds. It has more predictive power and can effectively estimate body weight in any species (Mavule et al., 2013) as it uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (Jolliffe, 2002). PCA of body measurements has been used as a tool in breed description and characterization of different sheep breeds (Cerqueira et al., 2011; Legaz et al., 2011;) as well as to predict body weight (Mavule et al., 2013; Eyduran et al., 2013). Nellore is an important sheep breed of South India (Kamalakar et al., 2017). Use of the principal component scores in linear regression for predicting body weight in extensively raised Nellore Palla sheep remains unexplored. Hence, the present study was carried out to establish the relationship between body measurements and body weight, and to evaluate the efficacy of prediction models derived from original measurements versus principal component scores.

Materials and Methods

The present study was conducted in SPSR Nellore which is the southernmost district of one of the nine erstwhile coastal districts of Andhra Pradesh, India. It lies between 14˚ 6' and 15˚ 51' N of the Northern Latitude and 79˚ 0' and 80˚ 42' of the Eastern Longitude spreading over an area of 10,440 sq.kms accounting for about 6.40 percent of total area of the state. The climate in the home tract is generally dry and it lies on the area of precarious with uncertain rainfall. Generally, April, May and June are the hottest months with highest temperature recorded as 46.7°C whereas, the temperature is low in the month of December and January with minimum temperature recorded as 22.8°C) and the average rainfall of the district is 1052.9mm. A comprehensive set of body measurements encompassing linear dimensions (height at withers, body length) and circumferential measurements (chest girth, paunch girth) and head and appendage measurements (face length and width, horn length, tail length and ear length) were recorded from 560 Nellore Palla sheep (489 ewes and 71 rams) during the year 2023-2024. Age of the animal was assessed through the time-honoured method of dental examination, with only those sheep exhibiting two or more permanent incisors, signifying adulthood are included in the analysis. To mitigate potential errors from gut fill, all measurements were precisely recorded during the early morning hours with the animals standing in their natural posture, utilizing standard measurement tape for body measurements with the nearest accuracy of 0.1cm, and weighing balance for body weight (kg).

Statistical analysis

Statistical analysis was conducted using SPSS v 20.0 and analysis of variance was performed to assess the effect of age and sex on the morphometric traits. Pearson's correlation coefficients were computed to examine the linear relationship between body measurements and body weight, and their statistical significance was determined. The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy and communalities was derived to validate the application of PCA, using a KMO threshold of 0.50. Bartlett's sphericity test was used to ensure that the correlation matrix was either the identity matrix or not. Principal component analysis, employing the Kaiser criterion (Eigen values >1) and Varimax rotation was used to enhance the interpretability of the PCs. Stepwise multiple regression was applied to develop models for predicting body weight from body measurements and PCA-derived components.
The following regression models were used

$$Y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\beta_3x_{i3}+\ldots\ldots\ \beta_px_{ip}+\varepsilon_i,\ \text{for } i=_i1,2,3,\ldots..n \quad (1)$$

$$Y_i=\beta_0+\beta_1PC_{i1}+\beta_2PC_{i2}+\beta_3PC_{i3}+\ldots\ldots\ \beta_pPC_{ip}+\varepsilon_i,\ \text{for } i=_i1,2,3,\ldots,n \quad (2)$$

where Y_i is the value of the i^{th} observation;

β_0 is the intercept; β_1 β_2 β_3......... $\beta_p$are the p^{th} partial regression coefficients; x_{i1}, x_{i2} ,x_{i3}..... x_{ip}; PC_{i1}, PC_{i2}, PC_{i3}......... PC_{ip}, are the p^{th} original body measurements and principal component scores, respectively, for the i^{th} observation; ε_i is the residual error, assumed as statistically independent, with common mean 0 and variance $\sigma^2$and are approximately normal in distribution.

The accuracy of prediction models were evaluated using the coefficient of determination (R^2) and the root mean squared error (RMSE). Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were used to evaluate model quality, specifically goodness of fit and model complexity (Akaike, 1974; Schwarz, 1978).

Results and Discussion

Descriptive statistics for live weight and body measurements are detailed in Table 1. The data revealed significant phenotypic variability in body weight (BW), face width (FW), and tail length (TL), with coefficients of variation between 13.19% and 14.50%. In contrast, other traits exhibited lower variability (CV 5-8%). The average body weight was 35.9 ± 0.20 kg, while average body length, height at withers, chest girth, and paunch girth were 70.2, 77.8, 82.4, and 81.1 cms respectively. These statistics are essential for informed management decisions, leading to optimal production efficiency.

Correlation Analysis

Pearson's coefficients of correlation (r) obtained between live weight and body measurements, and among body measurements are presented in Table 2. All the body measurements except TL were positively and significantly (0.227 to 0.938; P < 0.05/0.01) correlated with BW (0.009, P > 0.05). Chest girth had the highest correlation coefficient (0.938; P < 0.01) with Body Weight, followed by Height at Withers (0.916) and Paunch Girth (0.883). However, ear length, face length, face width combinations have significant but lower correlation coefficients with body weight and other morphometric traits. Since, the correlation between tail length and other traits is non-significant it is removed from the further downstream analysis. The correlation coefficients found in this study are higher than those reported by Mavule et al. (2013) in Zulu sheep breed. The observed high and positive correlations between body weight (BW) and morphometric traits suggests an increment in any body measurement results in a concomitant increase in live body weight. Whereas, the inclusion of low correlated traits such as ear length, face length, face width in selection criteria will not yield significant phenotypic gains in body weight (BW) or other morphometric traits of economic importance. The trait relationships identified in this study offer potential selection criteria, tailored to meet the production goals, breeding objectives, and socio-cultural considerations relevant to the farming community (Birteeb et al., 2012). Among all the body measurements studied, chest girth showed the highest correlation with body weight(r=0.938) and emerged as the most reliable predictor of body weight on Nellore Palla sheep.

Principal Component Analysis (PCA)

The high Kaiser-Meyer-Olkin (KMO) value of 0.850 for Palla sheep confirmed the suitability of the correlation matrix for principal component analysis (PCA), suggesting the presence of meaningful principal components. Furthermore, Bartlett's sphericity test yielded a highly significant chi-square statistic ($\chi^2 = 4402.598$, P<0.001), reinforcing the appropriateness of PCA.

Table 1. Descriptive statistics, sexual dimorphism of morphometric traits in Palla strain of Nellore sheep

Traits	Statistics		Effects		Sex differences		Sexual dimorphism
	Mean ± S.E	CV (%)	Age	Sex	Male (71)	Female (489)	
BL	70.21±0.17	5.94	***	***	74.16±0.51	69.64±0.17	1.06
HW	77.8±0.1	6.05	***	***	81.56±0.55	77.33±0.20	1.05
CG	82.4±0.17	5.09	***	***	84.98±0.61	82.10±0.17	1.03
PG	81.19±0.17	5.15	***	***	83.43±0.56	80.86±0.18	1.03
FL	18.57±0.06	8.02	*	*	18.70±0.18	18.55±0.06	1.007
FW	9.64±0.05	13.19	NS	NS	9.81±0.15	9.62±0.05	1.02
EL	15.62±0.05	7.81	NS	NS	15.73±0.16	15.60±0.05	1.007
TL	10.03±0.06	14.50	NS	NS	10.12±0.14	10.02±0.06	1.01
BW	35.90±0.20	13.73	***	***	40.28±0.61	35.26±0.20	1.14

Note: Body Length (BL), Height at Withers (HW), Chest Girth (CG), Paunch Girth (PG), Face Length (FL), Face Width (FW), , Tail Length (TL), Body Weight (BW). *** indicates P<0.001; * indicates P<0.05; NS: Non-significant.

Table 2. Correlation coefficients among morphometric measurements in Palla sheep

	CG	BL	PG	HW	FL	FW	EL	TL	BW
CG	1	0.811**	0.918**	0.883**	0.303**	0.245**	0.285**	0.008	0.938**
BL		1	0.786**	0.877**	0.252**	0.194**	0.226**	-0.004	0.797**
PG				0.863**	0.300**	0.229**	0.259**	-0.005	0.883**
HW				1	0.297**	0.220**	0.263**	0.006	0.916**
FL					1	0.430**	0.298**	0.177**	0.314**
FW						1	0.226**	0.243**	0.227**
EL							1	-0.068	0.266**
TL								1	0.009
BW									1

Note: Body Length (BL), Height at Withers (HW), Chest Girth (CG), Paunch Girth (PG), Face Length (FL), Face Width (FW), Horn Length (HL), Tail Length (TL), Body Weight (BW). *** indicates P<0.001; * indicates P<0.05; NS: Non-significant.

Eigen values and their share of total variance along with factor loadings measured traits are summarized in Table 3. Following Keiser criterion (eigen value>1), two factors (PC1 and PC2) were extracted as shown in the form of scree plot in Figure 1. Of which, the first principal components (PC1) explained 55.99 percent of total variation, provided effective description of body size defined by high positive loadings for body weight (BW), body length (BL), height at withers (HW), chest girth (CG), and paunch girth (PG). PC2, explained 19 percent of variance, described body shape and was defined by high positive loadings for face length (FL), face width (FW), and ear length (EL). Collectively, the two Principal components are explaining 76.55 percent of total variance of the population. The remaining unexplained variation may be due to segregation of casual alleles at contributory loci, measurement error and environmental factors (Mishra et al., 2017). The reduced dimensionality resulting from the two-PC extraction enables improved interpretability of trait intercorrelations and facilitates the application of more parsimonious model specifications. (Mota et al., 2016). Despite the orthogonality of principal components, the traits loading on one component are independent of those loading on other components. So, selection to improve body size, an important target for meat production (Eyduran et al., 2013) implies little variation in body shape. Multi-dimensionality scaling revealed that all the traits were categorised into two groups ascribing to principal components explaining body size and body shape respectively as shown in Figure 2.

Parallel to the present study, Mavule et al. (2013) have documented comparable results, characterizing the first factor as indicative of general body size. The percentage of variance explained by PCA in this study was greater than that reported for Zulu sheep (Mavule et al., 2013) and Madgyal sheep (Yadav et al., 2016) and lower than that for Djallonke sheep (Birteeb et al., 2014) and Kajali sheep (Mishra et al., 2017) which revealed a notable level of conformational variation. Similar to the present study, Mavule et al. (2013) and Cerqueira et al. (2011) also reported two PCs are sufficient to explain most of the total variation in young Zulu sheep and Portuguese Bordaleira sheep populations respectively. It is clear that variability of body measurements might be different across breeds, but some traits commonly influence body weight in sheep.

Table 3. Varimax rotated factor loadings, eigen values, percentage of total variance and communalities of morphometric traits of Palla sheep

Trait	PC1	PC2	Communality	
			Extraction	Unique factors
Chest Girth	0.941	0.207	0.928	0.072
Body Length	0.893	0.133	0.815	0.185
Paunch Girth	0.922	0.191	0.886	0.114
Height at Withers	0.944	0.177	0.922	0.078
Body Weight	0.940	0.195	0.921	0.079
Face Length	0.164	0.789	0.65	0.35
Face Width	0.077	0.783	0.619	0.381
Ear Length	0.178	0.594	0.384	0.616
Eigen Values	4.794	1.330		
Total Variance	59.923	16.630		

Table 4. Step-wise multiple regression among morphometric traits of Palla Sheep

S. No	Models	Model equation	R^2	RMSE	AIC	BIC
Multiple Linear Regression						
1	CG	-54.77 + 1.1CG	87.9	1.72	263.37	264.12
2	CG, BL	-55.243 + 0.998CG +0.126BL	88.3	1.69	257.40	258.89
3	CG, BL, HW	-52.274 + 0.709CG - 0.147BL + 0.514HW	**91.7**	**1.416**	**173.51**	**175.76**
4	CG, BL, HW, PG	-52.358 + 0.7CG - 0.147BL + 0.512HW + 0.13PG	91.7	1.417	175.46	178.46
5	CG, BL, HW, PG, FL	- 52.929 + 0.697CG - 0.146BL + 0.509HW + 0.011PG + 0.058FL	91.7	1.416	176.63	180.37
6	CG, BL, HW, PG, FL, FW	- 52.914 + 0.7CG - 0.146BL + 0.508HW + 0.011PG + 0.078FL - 0.057FW	91.7	1.416	178.12	182.60
7	CG, BL, HW, PG, FL, FW, EL	- 52.562 + 0.70CG - 0.146BL + 0.509HW + 0.0097PG + 0.0855FL - 0.053FW - 0.049EL	91.7	1.416	179.73	184.97
Principle Components						
1	PC1	40.084+6.287PC1	92.3	1.61	175.84	176.59
2	PC1, PC2	38.928+5.833PC1+3.645PC2	**94.4**	**1.25**	**168.36**	**169.86**

Communalities quantify the proportion of variance in the original variables attributable to the extracted principal components. The communalities for Palla sheep traits, detailed in Table 3, were largely high, ranging from 0.384 (EL) to 0.928 (CG) and the unique factors ranged from 0.072 to 0.616 for all the traits under study which shows that most of the variances are shared between the variables allowing the use of PCA to classify them.

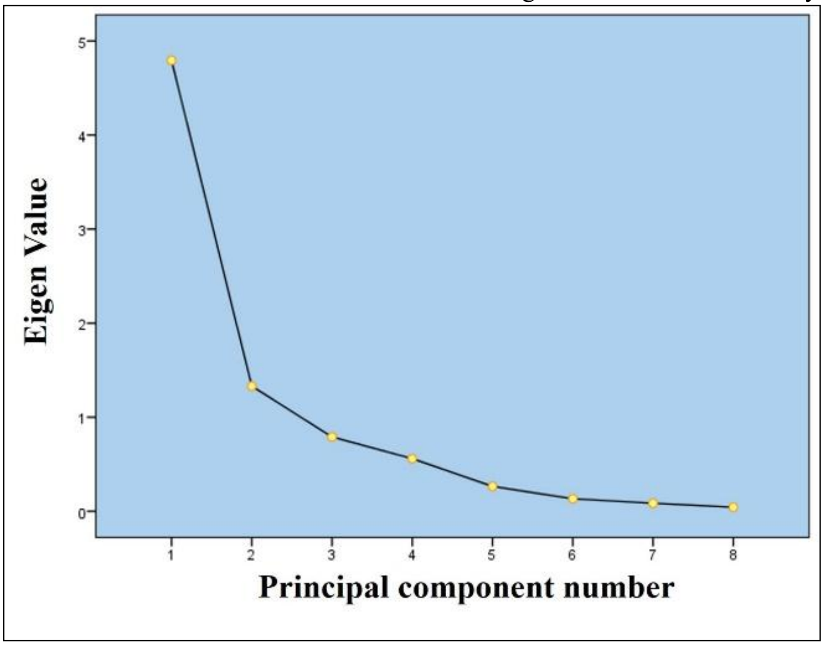


Figure 1: Scree plot showing the number of components extracted

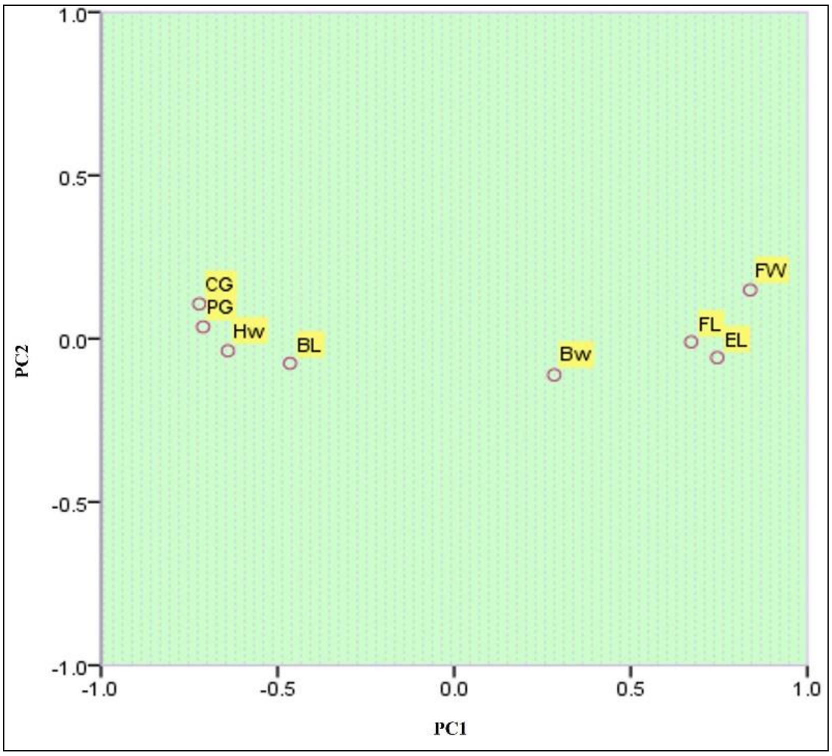


Figure 2: Score plot of the first two principal components of the body measurements of Palla sheep

(Note: BW: Body Weight; HW: Height at Withers; BL: Body Length, CG: Chest Girth, PG: Paunch Girth, FL: Face Length, FW: Face Width, EL: Ear Length)

a) Body length

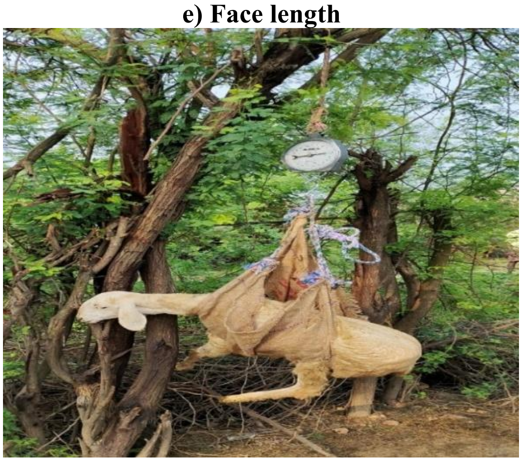
b) Height at withers

c) Chest girth

d) Paunch girth

e) Face length

f) Ear length



g) Weighing of an ewe

h) Weighing of a Ram

Figure 3(a-h) - Morphometric measurements and body weights of Palla strain of Nellore sheep

Prediction models for body Weight

The regression equations predicting body weight of Palla sheep from original body measurements along with PC scores are shown in Table 4. Results of the stepwise multiple regression analysis revealed that chest girth alone accounted for 87.9% of the variation in live weight. The inclusion of body length increased this proportion upto 88.3%. The accuracy of the model was further improved to 91.7% when height at withers were included in the equation. Besides the higher R^2, this model showed lowest AIC, BIC and RMSE values, which confirms its better goodness of fit when compared to the models that use original body measurements. Kunene et al. (2009) reported that linear regression coefficients of heart girth and wither height could be used to estimate the body weight of the Zulu sheep with R^2 of 0.66 to 0.49, respectively. In the Madgyal sheep, Yadav et al., 2016 reported highest R^2 value (0.886) with the inclusion of chest girth, body length, height at withers, thoracic depth and Shin circumference to predict body weight.

It was observed that goodness of fit increased by including the two PCs in the model and it has demonstrated a strong explanatory power (R^2), accounting for 94.4% of the body weight variation with lower root mean squared error (RMSE), Akaike's information criterion (AIC), and Bayesian information criterion (BIC). While adhering to the principle of parsimony, this two-PC model was more efficient than the best-performing model using original measurements, which required three variables. In a similar study, Yadav et al. (2016) have found that body weight prediction model with PC1 and PC2 presented better R^2 (0.94) and lowest RMSE (1.86) rather the model with only PC1 ($R^2 = 0.80$ and RMSE=3.84) in a Madgyal sheep population. The PCs obtained in this study can be utilized as a composite variable set or within a selection index, reducing computational demands through the application of two weighted coefficients (Mota et al., 2016).

Effective sheep breeding programs prioritize increase in body weight, necessitating a comprehensive understanding of correlated traits to inform selection strategies (Mavule et al. 2013). The reliability of chest girth as a sole predictor of body weight has been established in multiple studies, due to the observed high associated regression coefficients. (Kunene et al. 2009). The phenomenon of multicollinearity among predictor variables can induce instability in the estimation of regression coefficients. The application of principal component analysis (PCA) factor scores, which are by definition orthogonal and therefore uncorrelated, resolves the issue of collinearity. Hence, utilization of principal component scores may provide a more robust prediction of body weight than the original inter-correlated variables

Conclusion

The findings of this research demonstrate the feasibility of predicting Palla sheep live weight through body measurements, providing a practical alternative for farmers who do not have access to weighing scales. Additionally, the development of a prediction model using two principal components, derived from the body measurements, resulted in improved predictive accuracy compared to models using the original variables. These two principal components exhibit a strong association with the most pertinent body measurements, indicating their effectiveness in capturing the essential information.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1) Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control ,19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

2) Birteeb, P. T., Peters, S.O., Yakubu, A., Adeleke, M. A. and Ozoje, M. O., 2013. Multivariate characterisation of the phenotypic traits of Djallonke and Sahel sheep in Northern Ghana. Tropical Animal Health and Production, 45(1), 267–74.

3) Cerqueira, J.O.L., Feás, X., Iglesias, A., Pacheco, L.F. and Araújo, J.P.P., 2011. Morphological traits in Portuguese Bordaleira de Entre Douro e Minho sheep: divergence of the breed. Animal Production Science, 51(7),635–641. https://doi.org/10.1071/AN10147

4) Eyduran, E., Waheed, A., Tariq, M.M., Iqbal, F., Ahmad, S.,2013. Prediction of live weight from morphological characteristics of commercial goat in Pakistan using factor and principal component scores in multiple linear regression. Journal of Animal Plant Sciences. 23,1532–1540.

5) Jolliffe, I.T., 2002. Principal Component Analysis. Springer, 2nd ed., New York.

6) Kamalakar, G., Rambabu, K., Sumiran, N. 2017. Surgical management of traumatic intestinal eventration in a Nellore Brown Ewe. Journal of Livestock Science 8: 107-108

7) Kunene, N.W., Nesamvuni, A.E. and Nsahlai, I.V., 2009. Determination of prediction equations for estimating body weight of Zulu (Nguni) sheep. Small Ruminant Research, 84(1-3), 41–46. https://doi.org/10.1016/j.smallrumres.2009.05.003

8) Legaz, E., Cervantes, I., Pérez-Cabal, M.A., Fuente, L.F., Mártinez, R., Goyache, F., Gutiérrez, J.P.,2011. Multivariate characterisation of morphological traits in Assaf sheep. Small Ruminant Research, 100, 122–130. https://doi. org/10.1016/j.smallrumres.2011.06.005
9) Mavule, B.S., Muchenje, V., Bezuidenhout, C.C., Kunene, N.W., 2013. Morphological structure of Zulu sheep based on principal component analysis of body measurements. Small Ruminant Research, 111(1), 23–30. https://doi.org/10.1016/j. smallrumres.2012.09.008
10) Mishra, A.K., Vohra, V., Raja, K. N., Singh, S. and Singh, Y., 2017. Principal component analysis of biometric traits to explain body conformation in Kajali sheep of Punjab, India. Indian Journal of Animal Sciences, 87(1), 93-98. 10.56093/ijans.v87i1.66914.
11) Mota, R.R., Costa, E.V., Lopes, P.S., Nascimento, M., Silva, L.P., Silva, F.F., Marques, L.F.A., 2016. Multi-Trait analysis of growth traits: fitting reduced rank models using principal components for Simmental beef cattle. Ciencia Rural, 46: 1656–1661. https://doi.org/10.1590/0103-8478cr20150927
12) SAS Institute Inc., 2003. SAS User's Guide: Version 9.1. Cary, North Carolina:SAS Institute Inc.
13) Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics, 6(2),461–464. https://doi.org/10.1214/aos/1176344136
14) Yadav, D.K., Arora, R., and Jain, A., 2016. Principal component analysis of body measurements based morphological structure of Madgyal sheep. Indian Journal of Animal Sciences, 86(5), 568-571. 10.56093/ijans.v86i5.58497.