International Journal of Veterinary Sciences and Animal Husbandry

ISSN: 2456-2912
VET 2024; 9(3): 264-268

www.veterinarypaper.com
Received: 25-02-2024
Accepted: 29-03-2024

Dimpi Khanikar
Ph.D. Scholar, Department of Animal Genetics and Breeding, College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati, Assam, India

Arundhati Phookan
Assistant Professor, Department of Animal Genetics and Breeding, College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati, Assam, India

Eyangshuman Das
MVSc Scholar, Department of Animal Genetics and Breeding, College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati, Assam, India

TC Tolenkhomba
Professor, Department of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Central Agricultural University, Selesih, Aizawl, Mizoram, India

Corresponding Author:
Arundhati Phookan
Assistant Professor, Department of Animal Genetics and Breeding, College of Veterinary Science, Assam Agricultural University, Khanapara, Guwahati, Assam, India

Principal component analysis and its role in animal science: A review

Dimpi Khanikar, Arundhati Phookan, Eyangshuman Das and TC Tolenkhomba

Abstract
Principal component analysis (PCA) reduces an original group of variables into another smaller group called principal components which basically are the linear combination of original variables. The purpose of this multi factorial analysis is to reduce a set of data that may describe the subject and be used easily. PCA in the studies of animal science has played a pivotal role. PCA has been used not only for assessment of the body shapes but also as permit an understanding of the complex growth process going on in the body dimensions of an animal during growth. Results of PCA has been impactful in the management of animals, their conservation and selection of multiple traits by breeders.

Keywords: Eigen values, correlation, inter-correlated variables, orthogonal, principal component

Introduction

Principal Component Analysis, the oldest of multivariate techniques can be dated back to the scientific studies of Pearson in the 1980s or even Cauchy in 1901. However, it was formally introduced with its current terminology "principal components" by Hotelling (1933) [6]. The technique comprises of analyzing an extensively detailed table in which each observations are described by several inter-correlated quantitative dependent variables. This is followed by the extraction of relevant information from the table to represent it as a set of new orthogonal variables called principal components. Further, this technique reflects the pattern of similarity among observations and variables in the form of points in maps. A few principal components are responsible represent relationship among sets of many interrelated variables. As for instance, analysis of variance and correlations are widely used to characterise phenotypic and genetic relationships among productive or reproductive traits of animals. However, PCA allows the use of fewer measurements without significant loss of information.

As PCA analyses a data set it mainly extracts the most pertinent information from the data set. It then untangles the description from the reduced data set. Followed by which it scrutinizes the structure of the observations and the variables. In order to achieve this target, the new variables designated as principal components are obtained as linear combinations of the original variables which represent the most relevant information. The first principal component is required to have the largest possible variance (Meaning, will represent the largest part of the data set). The second component is calculated since it is orthogonal to the first component. The other components are extracted in the similar manner (Abdi and Williams, 2010) [1]. Although the number of components generated in PCA equals the number of variables in the analysis, first few components account for the highest proportion of the total variance (Yunusa *et al*., 2013) [25].

PCA as a medium for simpler and efficient study in animal science

Following are few of the studies under various section heads where PCA has proved its efficiency in various assessments.

Morphology and Biometric traits assessment: The different body measurements representing the size of the animal is an important criterion for selection of animal.

Body measurements and indices estimated from different combinations of different body traits can also be used as an indicator of type and function in domestic animals (Schwabe and Hall, 1989; Salako, 2006) [18, 17]. As for example, Tolenkhomba *et al.* (2013) [21] applied PCA to study the interdependence among five biometric traits *viz*. body length (BL), head length (HL), chest girth (CG), front height (FH) and rear height (RH) in Mizo local pigs (Zovawk) at different stages of life. A single factor extracted each at birth, 4 week, weaning (52 days), 14 and 18 weeks of age, accounted for 60.41%, 69.79%, 67.49%, 81.85% and 87.73% of the total variance, respectively. Therefore, the study proved that principal components can prove to be helpful in breeding and selection programmes in order to achieve coordinated livestock morphometry using lesser measurements. The exploitation of body dimensions could be achieved when they are grouped more meaningfully. Mishra *et al.* (2017) [8] conducted a study where different body biometric traits were analysed in Kajali ewes of Punjab (India) using PCA with Kaiser Normalization to explain their body conformation and were subsequently used to predict adult body weight. Three major principal components were extracted which explained 68.66% of the total variation regarding body biometry. First component illustrated the body size and explained about 36% of total variation. The second component described the tail length, height and ear length and explained about 21% variation. The study showed that compared to multiple regression analysis PCA is more appropriate for predicting adult body weight. Putra and Ilham (2019) [12] aimed to investigate the body measurements and body indices of Katjang does of Indonesia. PCA was conducted on 11 body measurements of face length (FL), face width (FW), face height (FH), ear width (EW), ear length (EL), chest width (CW), chest depth (CD), chest girth (CG), cannon bone circumference (CC) and eleven body indices of cephalic index (CpI) length index (LI), depth index (DI), body length (BL), body index (BI), withers height (WH), conformation index (CI), proportionality (Pr), relative depth of thorax (RDT), dactyl thorax index (DTI), thoracic development (TD), area index (AI), and relative cannon thickness index (RCTI) were calculated in this study. As first component (PC1) four components of body measurements (FH, BL, CD, CG) and three components of body indices (CI, TD, RCTI) were identified for Katjang does. The result suggests that principal components were appropriate predictor of biometrics of Katjang does. In order to define breed standards and to understand developmental ability of the animals linear type traits are important. For data analysis, PCA is the most efficient technique when variables are correlated. On this note Dahiya *et al.* (2020) [4] conducted a study with a motive to make linear type traits unrelated and reduce their number to the extent which could be used in explaining body conformation in Murrah buffaloes. Measurements were recorded from 81 adult Murrah buffaloes for 11 linear type traits which includes top wedge angle, rump slope, rump width, hip bone distance, navel flap length, brisket distance, height at wither, body length, skin thickness at neck region, skin thickness at ribs region and skin thickness at rump region. PCA analysis revealed four components which detailed total variation of 69.522% out of which first component described 28.678% variation. Thus PCA can be efficiently used in lessening the number of variables needed for explaining the body conformation in Murrah buffaloes. Similarly, Putra *et al.* (2020) [14] conducted a study with a data set of 7 body measurements and 16 body indices parameters from 144 heads of Pasundan cow West Java Province of Indonesia. The body measurements in this study consisted of body length (BL), withers height (WH), chest girth (CG), rump length (RL), chest width (CW), rump height (RH) and rump width (RW). The body indices included width slope (WS), height slope (HS), area index (AI), body index (BI), conformation index (CI), rump length index (RLI), length index (LI), proportionality (Pr), body ratio (BR), thoracic development (TD), transverse pelvic (TP), pelvic index (PI), and longitudinal pelvic (LP). The PCA revealed two factors of PC1 (WH, BL, RH) and PC2 (CG, CW, RW, RL) out of 73.36% of the total variation. PCA for body indices was observed four factors of PC1 (BI, AI, RLI, LI, Pr, LP), PC2 (CI, TD), PC3 (WS, PI, TP) and PC4 (HS, BR) out of 89.38% of the total variation. Hence, the seven body measurements are important for detailing the body conformation of Pasundan cows which are body size (PC1) and body shape (PC2). In order to explain the relationship between morphologic traits of Large-white and Duroc pig breeds Akporhuarho and Iriakpe (2021) [2] used to identify components that best define body conformation. Body weight along with 5 biometric variables *viz*. head length, body length, body girth, ham length and ear length were included. Two principal components were identified for Large-white whereas there were three components of Duroc. The first factor for both breeds having the largest percentage of the total variation designated the general size. The variation for body weight in Large-white and Duroc respectively was accounted for 58 and 76% the study concluded that the use of PCA techniques allows to explore the interdependence in the original five parameters measured: head length, body length, body girth, ham length and ear length of Large-white and Duroc. Tolenkhomba *et al.* (2021) [20] used PCA s an attempt to know about morphological structures of local bulls "Zobawng" of Mizoram for breed characterization. Different biometric traits including body length, heart girth, paunch girth, height at wither, forehead width, elbow length, arm length, fore-shank length, hind shank length, thigh length, tail length, ear length, head length, switch length, eye to eye space, neck length, circumference of horn, circumference of neck, and space between horns were taken into account in 50 bulls of Mizoram local cattle 'Zobawng' were recorded and analyzed by PCA to explain body conformation. Five factors were highlighted using PCA analysis with promax rotation which explained 78.37% of the total variation. Factor 1 represented the general body conformation (45.56% of total variation). The objective of research by Misrianti *et al.* (2023) [9] was to characterized morphology and estimated genetic distance between intra population of Kuantan cattle. Five variables were measured: body length (BL)(cm), hip height (HH) (cm), wither height (WH)(cm), chest girth (CG)(cm), and chest depth (CD)(cm). The first factor in PCA representing body measurement contributed 32.77%, and the second factor representing body shape contributed 25.83% of total variability. The dendrogram showed there are three clusters of Kuantan Cattle, thus proving the efficiency of PCA. Putra *et al.* (2021) [13] aimed to characterize Thin-tail sheep in the highland and lowland areas of Jambi Province using multivariate analysis and taking into account 7 body measurements and body indices. Out of the total phenotypic variance of animals, PCA explained 65.84%-72.30% by body measurements and 78.23%-84.99% by body indices. The cluster analysis of the sheep population revealed cluster 1 (Kerinci and Sungai Penuh) and cluster 2 (Muaro Jambi and Batanghari). It was concluded with the aid of PCA about 60%

of Thin-tail sheep could be characterized through their body indices. In another similar study, Yakubu *et al.* (2022) [24] in case of Nigerian indigenous pigs, aimed at understanding the interdependence among the morphological and heat tolerance traits as well as to predict body weight from their conformation traits. Data on body weight, eight linear body measurements and three thermo-physiological parameters were measured on selected pigs of three growth stages (piglets, growers and finishers). To determine size, shape and heat multivariate PCA was used. Three principal components (PC1, PC2 and PC3) namely height at withers, ear length and body weight were found to be the most discriminating variables to distinguish the pig. It was shown henceforth that 100% of piglets, 96.7% of growers and 96.7% of finishers were correctly categorized in their distinct populations. A study by Warman *et al.* (2023) [23] aimed to provide information about female Bali cattle's morphometric characterization and zoometric indices using PCA. Morphometric characteristics in this study consists of body length (BL), chest girth (CG), withers height (WH), chest depth (CD), chest width (CW), rump width (RW), rump height (RH), head width (HW) and head length (HW). Therefore, the zoometric indices included body index (BI), cephalic index (CeI),body ratio (BR), compact index (CoI), depth index (DI), index of compression (IC), conformation index (CnI), height index (HI), height slope (HS) over increase index (OII),), length index (LI), weight (W). thoracic development (TD) and transverse pelvic (TP). The PCA for morphometric characteristics revealed two factors of PC1 (BL, CG, CW, CD, RW, and HL) and PC2 (WH, RH, and HW), which explain about 67.50% of the total variation. PCA for zoometric indices sorted out four factors of PC1 (CoI, CnI, DI, LI, TD, TP, W), PC2 (BR, HS, and OII), PC3 (BI, IC), and PC4 (CeI), which explained 90.64% of the total variation. It was concluded that BL, CG, CW, CD, RW, and HL were important to describe the body conformation of female Bali cattle and could be used as selection criteria. Patel and Rank (2023) [11] attempted to define the biometric traits of HF crossbred cattle from three farms around Anand (India). The biometric traits were age, live body weight (BW), body length (BL), height at hip (HH), height at wither (HW), heart girth (HG) width of hip (WH) and chest depth (CD) out of which PCA revealed two principal components each were extracted in female and male group. The identified two components in female group could explain 95.88% of cumulative variance. First component accounted for 69.20% of the variation. The second component explained 26.68% of total variance. Out of two principal components, one aided in the reduction in the number of biometric traits to be recorded in HF crossbred female. In HF crossbred male group, two components were identified that could explain 97.57% of cumulative variance. First and second component explained 60.58% and 36.99%, respectively of total variance. First component represented maximum of general body conformation in HF crossbred male. PCA proved its efficiency in breeding programs with a significant reduction in the number of biometric traits so as to explain body conformation and hencefoth selection of superior animals. Thobela and Putra (2023) [19] aimed to perform PCA in the body measurements (face width, face length, body length, ear length, rump height, sternum height, withers height, rump width and chest girth) of adult Nguni cattle. Three principal components (PC) were obtained that explained about 68.12% of total variance. The first component (PC1) explained the morphostructure of the animal (Face width, body length, rump height and withers height) for about 24.31%. In conclusion, the results of PCA in was found to be accurate and could be used for selection program of Nguni cows.

- **Assessment of reproductive traits:** Ali *et al.* (2013) [3] conducted a study to assess more features for various productive and reproductive traits accumulated over 10 and 13 years at a livestock experiment station, Rakh Gulama, Bhakkar, Pakistan which has data on 74 Holstein-Friesian x Sahiwal crossbred and 237 Nili-Ravi buffaloes, respectively. The first principal component (PC1) elucidates 41.8% of the total variation for different traits of buffalo data while it is 44.78% for different traits of cattle data. The second, third and fourth principal components for different traits of buffaloes explains 26.33%, 17.67% and 12.44% of the total variation respectively while those of 24.11%, 18.22% and 11.44% for various traits of cattle data. On the basis of these results it was suggested that the variables under study can be classified into three clusters i.e. (a) dry period, calving interval, service period and overall average (b) age at first calving and dry period, and (c) milk yield, wet average and lacation period to mitigate the misleading results that are revealed using univariate techniques. Panda *et al.* (2020) [10] undertook an investigation to derive fewer independent reproductive traits through PCA by using records of 2 litter traits at birth, 4, 6 and 8 week from 42 crossbred (75% Landrace X 25% Bareilly local) pigs. A single principal component (PC) was extracted for litter traits which accounted for 82.24% of total variance. PC1 comprised of litter size at 4, 6 and 8 weeks of age. This factor seemed to be representing the overall performance of the litter which therefore proved that PC may be exploited in breeding and selection program for reproductive traits. Sarma *et al.* (2022) [16] focused his study to identify the principal components for economic traits in layer chicken from a data set which included weekly body weight (g) from 0 day to 20th and 40th week, age at sexual maturity (days), body weight (g) at first egg production, weight of first egg (g), egg weight at 40th week (g), egg numbers at 40th week, egg numbers at 52nd week, egg weight at 52nd week (g). Eventually, a total of three principal components were obtained which explained a total variance of 75.524%. Principal component 1 had high loads on body weight 10th week to 20th week and had a variance of 38.892%. Similarly, PC2 and PC3 explained variance of 27.072% and 9.560% respectively and had high loads on 1 week body weight to 9-week body weight (BW1-BW9) and age at sexual maturity, 40-week egg production, 52-week egg production respectively. The study proved that PCA can be a tool for selecting the economic traits for breeding of layer chicken**.**
- **Meat quality assessment:** Karlsson (1992) [7] applied PCA for evaluating results from pig meat quality measurements i.e. pH, meat colour, protein extractability and pigment content in Swedish pig carcasses The result indicated that when using PCA for selection among the meat quality methods used, the greatest proportion of the total variance was explained by the ultimate internal reflectance. The results of this study concluded that PCA is a simple method of finding objects with different characteristics and for variable selection. Ros-Freixedes *et al.* (2014) [15] used principal component analysis to determine the relationship between meat quality traits, feeding patterns, scale activity, and number of conflict–avoidance interactions. The first PC indicated that gilts with

greater daily feed intake stayed longer in the feeder and their meat had increased intramuscular fat (IMF), was lighter in color, and, in the second PC, had better juiciness, tenderness, chewiness, and flavor. The third PC suggested that dominant gilts could gain priority access to the feeder, eating more and growing fatter. Hence, with the help of PCA, it has been revealed that gilt scale activity and conflict-avoidance behaviors were not good indicators of final meat quality attributes.

Estimation of Genetic Parameters: Viana *et al*. (2020) [22] estimated genetic parameters for 7 traits of Nellore, to verify how the estimate breeding values (EBVs) of the traits are distributed in different Brazilian states, and to suggest a selection index by state/sex. Single-trait analysis was used to estimate heritability (h^2) and EBVs under animal model. PCA was used to explore relationships among animal EBVs for these traits. The first 2 principal components showed correlation above ±0.60 with EBVs of all analysed traits, and retaining above 96% of the total breeding value variance. Selection on the basis of PC1 could be of aid to identify animals with favorable breeding values among all studied traits. Therefore it was concluded that PCA is a good alternative in the elaboration of selection indices in Nellore breed defined for different sex and environments.

PCA as a tool for breed identification: The economic value of livestock lies in the appropriate identification of breeds. Owing to the large intra-breed variation traditional breed identification methods such as coat colour are prone to error. PCA, according to a study conducted by Dan *et al.* (2023) [5] can categorize pig breeds using their images. For this purpose individual images of five different pure breeds were captured under both controlled and uncontrolled environments from organized farms in India. Three different image sets (In the controlled, uncontrolled, and mixed environment) were created. The accuracy level was found to be at 93%. Finally, it was concluded that PCA method outperformed any other method of breed identification.

Conclusion
PCA showcases number of benefits. Few of them includes dimensionality reduction (helps in to visualize or analyze data), feature extraction (of features or elements from the original data that might be more insightful or understandable than the original features), data visualization (high-dimensional data in two or three dimensions), noise/error reduction and reducing the effects of multicollinearity (by identifying the most crucial features or components). Adjacently, PCA carries few disadvantages there is a difficulty or information loss upon dimensionality reduction, the covariance matrix can be distorted by outliers, which can make it harder to identify the most crucial characteristics and finally extremely huge data sets may prove be a difficulty in computation. In spite of the cons, PCA has become a boon to the field of research and shall prove to be an efficient aid in data analysis in the years to come.

Author's Contribution
Not available

Conflict of Interest
Not available

References

1. Abdi H, Williams LJ. Principal component analysis. Computational Statistics. John Wiley and Sons; c2010. p. 433-459.
2. Akporhuarho PO, Iriakpe O. Assessment of body weight and morphological traits of two breed of grower pigs using principal component analysis. Niger J Anim Prod. 2021;48(5):1-11.
3. Ali A, Ullah MA, Ejaz MA. Evaluation of livestock data using multivariate analysis techniques. Pak J Life Soc Sci. 2013;11(3):208-211.
4. Dahiya SP, Kumar M, Dhillod S, Ratwan P. Principal component analysis of linear type traits to explain body conformation in Murrah buffaloes. Indian J Anim Sci. 2020;90(11):1546-1550.
5. Dan S, Mandal S, Ghosh P, Mustafi S, Banik S. Principal component analysis in pig breeds identification. Indian J Anim Sci. 2023 Apr;93(4):401-405. doi: 10.56093/ijans.v93i04.131010.
6. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;25:417-41.
7. Karlsson A. The use of principal component analysis (PCA) for evaluating results from pig meat quality measurements. Meat Sci. 1992;31(4):423-433.
8. Mishra AK, Vohra V, Raja KN, Singh S, Singh Y. Principal component analysis of biometric traits to explain body conformation in Kajali sheep of Punjab, India. Indian J Anim Sci. 2017;87(1):93-98.
9. Misrianti R, Mainidar J, Asharudin HB, Dedi YS, Ali A, Wijaya SH, *et al*. Determination of morphological characteristics in Kuantan cattle using multivariate analysis. Bull Anim Sci. 2023;45(3):142-147.
10. Panda S, Gaur GK, Jena D, Chhotaray S, Tarang M, Wara AB. Principal component analysis of litter traits in crossbred piglets. Pharma Innov J. 2020;9:31-33.
11. Patel AJ, Rank DN. Principal component analysis of biometric traits explain the body weight of HF crossbred cattle. Indian J Vet Sci Biotechnol. 2023;19(2):34-39.
12. Putra WPB, Ilham F. Principal component analysis of body measurements and body indices and their correlation with body weight in Katjang does of Indonesia. J Dairy Vet Anim Res. 2019;8(3):124-134.
13. Putra WPB, Alwi Y, Suryani H. Morphometrics characterization of thin-tail sheep in lowland and highland areas. Trop Anim Sci J. 2021;44(4):386-398.
14. Putra WPB, Syahruddin SAİD, Arifin J. Principal component analysis (PCA) of body measurements and body indices in the Pasundan cows. Black Sea J Agric. 2020;3(1):49-55.
15. Ros-Freixedes R, Sadler LJ, Onteru SK, Smith RM, Young JM, Johnson AK, *et al*. Relationship between gilt behavior and meat quality using principal component analysis. Meat Sci. 2014;96(1):264-269.
16. Sarma O, Dubey PP, Dash SK, Sahoo SK, Malhotra P. Principal component analysis of different economic traits in layer chicken. Agric Sci Dig. 2022. doi: 10.18805/ag.D-5519.
17. Salako AE. Principal component factor analysis of the morphostructure of immature Uda sheep. Int J Morphol. 2006;24(4):571-4.
18. Schwabe AE, Hakk SJG. Distokia in nine British breeds of cattle and its relationship to the dimensions of the dam and calf. Vet Rec. 1989;125:636-639.

19. Thobela LT, Putra WPB. Principal component analysis (PCA) in the body measurements of Nguni cows. Pak J Zool. 2023;55(3):1469.
20. Tolenkhomba TC, Anal W, Singh NS, Chaudhury JK, Mayengbam P. Principal component analysis of body measurements of Zobawng bulls: A local hill cattle of Mizoram, India. Heart. 2021;112(1.03):6-47.
21. Tolenkhomba TC, Saikia P, Hmar L, Prava M, Singh NS. Principal component analysis of body measurements in Mizo local pigs. Indian J Vet Res. 2013;22(1):26-31.
22. Viana AFP, Rorato PRN, Mello FCB, Machado DS, Figueiredo AM, Bravo AP, *et al*. Principal component analysis of breeding values for growth, reproductive and visual score traits of Nellore cattle. Livest Sci. 2020;241:104262.
23. Warman AT, Fadhilah GT, Ibrahim A, Atmoko BA, Baliarti E, Panjono. Morphometric characterization and zoometric indices of female Bali cattle reared in Lombok Tengah District, West Nusa Tenggara, Indonesia. Biodiversitas. 2023;24(2):966-974.
24. Yakubu A, Durven G, Hagan J. Multivariate analysis of body weight, morphometric and thermo-physiological traits of indigenous pigs under tropical conditions. Genet Biodivers J. 2022;6(1):91-104.
25. Yunusa AJ, Salako AE, Oladejo OA. Principal component analysis of the morphostructure of Uda and Balami sheep of Nigeria. Int. J Res Agric Sci. 2013;1(3):45-51.

How to Cite This Article
Khanikar D, Phookan A, Das E, Tolenkhomba TC. Principal component analysis and its role in animal science: A review. International Journal of Veterinary Sciences and Animal Husbandry. 2024;9(3):264-268.