Explor Anim Med Res.
Vol. 15, Issue 2, 2025
DOI: 10.52635/eamr/15.1.241-247

ISSN 2277-470X (Print), ISSN 2319-247X (Online)
Website: www.animalmedicalresearch.org

Research Article

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS FOR EVALUATING FIRST LACTATION TRAITS IN CROSSBRED CATTLE

Simran Kaur[1]*, AK Ghosh[2], Olympica Sarma[2], RS Barwal[2]

Received 12 November 2024, revised 27 February 2025

ABSTRACT: This study aimed to reduce redundancy among first lactation traits and identify key factors influencing production and reproduction in crossbred cows. Measurements were taken on 529 crossbred cattle (62.5% HF×37.5% Sahiwal) over a 30-year period (1990-2019) for ten traits for first lactation (age at first calving, lactation length, milk yield for first lactation, milk yield up to 305 days, first service period, date of dry, calving interval for first lactation, peak yield, days to reach peak yield and milk yield for up to 300 days). For the traits under study, phenotypic correlations were found, and significant positive correlations varied between 0.11 to 0.81. To explain the production and reproduction of crossbred cows, varimax rotational PCA with Kaiser Normalization was applied to all ten first lactation variables. The principal component analysis yielded four components that explained 77.495% of total variance, with the first component accounting for 34.374%. The PCA results indicated that milk production traits (MY305D, MY300D, FLMY, PY) were strongly associated with Component 1, while reproductive traits (FCI, DDRY) influenced Component 2. Traits like AFC and LL defined Component 3, highlighting early maturity and lactation performance, and Component 4 reflected peak production and service period efficiency. The communality of ten distinct first lactation traits varied from 0.997 (first calving interval) to 0.470 (first service period). Hence, PCA was found to be effective in properly capturing performance patterns in crossbred cows while lowering the amount of variables required to explain reproductive and productive efficiency.

Keywords: Principal component analysis, Crossbred cattle, Production, Reproduction traits.

INTRODUCTION

Selection based on lifetime performance is not realistically practicable due to the lengthy generation interval, hence it is desired to select animals based on the performance of earlier lactations rather than attributes that appear later in life [1]. Similarly, the suitability of breeding animals in an organized herd is essentially decided by their productivity and reproductive efficiency. Due to their strong correlation, first lactation attributes are indicative of subsequent lactation outcome in both cattle and buffaloes [2]. In order to effectively plan selection programme and provide information about trait variations, it is important to measure the phenotypic traits in dairy cattle herds. Breeding techniques developed based on single-trait selection cause misconceptions about the true performance of the herd [3]. In animal breeding programs that seek to elevate population variability by producing male and female animals with higher average performance relative to previous generations through heterosis, multivariate studies are helpful in making the right judgments. Moreover, univariate analyses have limitations when it comes to evaluating the variables related to milk production and reproduction traits because they assess each variable separately. In contrast, multivariate analyses simultaneously assess a set of characteristics while considering correlations between variables into account, which allows for more accurate

[1]*Department of Animal Genetics and Breeding, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences, Hisar, Haryana, India.*
[2]*Department of Animal Genetics and Breeding, College of Veterinary and Animal Sciences, G B Pant University of Agriculture and Technology, Pantnagar, Uttarakhand, India.*
**Corresponding author. e-mail: sudansimran321@gmail.com*

interpretations of the information gleaned from a data set [4]. A multivariate technique, such as principal component analysis, could be used to determine the loadings or factors that explain the most variation in the data set over dependent variables [5]. A mathematical technique called Principal Component Analysis (PCA) is employed to transform an array of associated variables into a lesser number of uncorrelated variables, lowering dimensionality. PCA can efficiently analyse a complete data set that includes numerous production and reproduction parameters such as milk yield, age at first calving and calving interval etc. among others [6]. PCA can be used to decrease the number of associated variables into a smaller collection of independent variables known as Principal components (PC), limiting the quantity of information lost from the original data. This strategy creates orthogonal components by linearly combining the key variables based on their eigenvalues. Each PC reflects more variability than the one after it, and the eigenvalues are arranged from maximum to minimum [7]. A significant proportion of common variance is shared by the traits loaded on the same PC [8]. Being a descriptive tool, PCA does not require distributional assumptions, it is considered as a highly flexible exploratory technique that may be used with a variety of numerical data formats [9]. PCA is a suitable method for data analysis in breeding and genetics because it permits the use of variables that are not measured in the same units [10]. Breeders can employ principal component analysis techniques to create selection indices [11]. Additionally, PCA is utilized to estimate genetic parameters as well as to uncover the possible biological associations among variables that were typically not seen in the original data as well as to look into genetic relationships between traits [12]. PCA makes data exploration easy and helps identify outliers and important variables by identifying the variance main axis within a set of data [13]. Lactation traits include both production as well as reproduction traits therefore multivariate approach like PCA can simultaneously assess several traits to explain the variability and can identify key factors influencing production and reproduction in a better way.

MATERIALS AND METHODS

Using Pearson's correlation approach, phenotypic correlations among economic attributes were computed [6]. Following that, multivariate principal component analysis was performed on the strongly associated traits. PCA aims to keep a higher percentage of variance from the original set of variables while reducing the number of composite variables. The data were collected from the pedigree sheets with complete data maintained at organised cattle herd reared at the institutional farm of Govind Ballabh Pant University of Agriculture and Technology (GBPUAT), Pantnagar. Bartlett's test (1950) was first used to assess whether the dataset, which included 10 variables and 529 crossbred cows (62.5% HF×37.5% Sahiwal), was suitable for factor analysis, as advised by [14]. The reliability of dataset was further confirmed using the Kaiser-Meyer-Olkin (KMO) test at a significance threshold of 1% for sample adequacy. Only factors with eigenvalues greater than one were kept after applying the Kaiser rule criteria [15] to determine the number of factors. The efficacy of common factor model was assessed using Kaiser's measure of sampling adequacy (MSA); an MSA of less than 0.5 was deemed undesirable. According to [16], PCA uses a correlation matrix to transform a collection of p variables (X1, X2, …, Xn) into a new set (Y1, Y2, …, Yp). Each main component (Yi) is denoted by a linear combination of standard attributes (Xj). Yi = ai1X1 + ai2X2 +…+ ainXn, where the eigenvectors of the correlation matrix for the traits being examined were denoted by aij. Orthogonal rotation was utilized to maximize variance in the linear transformation of the factor pattern matrix in order to enhance comprehension. The factor module in SPSS 24 was used to carry out the principal component analysis.

RESULTS AND DISCUSSION

Four principal components were obtained by applying PCA to ten distinct production and reproduction traits in crossbred cattle. The Kaiser-Meyer-Olkin (KMO) approach provided a sample adequacy assessment of 0.583 (Table 1), taking into account eigen values greater than one. This score examines the appropriateness of the data for each factor in producing accurate PCA findings. A KMO-MSA score of at least 0.5 is required for effective PCA analysis to proceed [17,18]. A sample adequacy score of less than 0.5 is considered unsatisfactory [19]. The sum

Table 1. KMO and Bartlett's test.

Kaiser-Meyer-Olkin	
Measure of Sampling Adequacy.	0.583
Approx. Chi-Square	7723.339
Bartlett's Test of Sphericity	
Df	45
Sig.	0.000

Table 2. Total variance explained by different components in crossbred cattle.

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.614	36.141	36.141	3.614	36.141	36.141	3.437	34.374	34.374
2	2.006	20.056	56.197	2.006	20.056	56.197	1.939	19.389	53.763
3	1.071	10.711	66.908	1.071	10.711	66.908	1.262	12.620	66.382
4	1.059	10.587	77.495	1.059	10.587	77.495	1.111	11.113	77.495
5	0.955	9.552	87.047						
6	0.817	8.169	95.216						
7	0.427	4.274	99.490						
8	0.046	0.459	99.949						
9	0.004	0.044	99.993						
10	0.001	0.007	100.000						

Table 3. Communalities for various traits in crossbred cattle.

Traits	Extraction
AFC	0.536
LL	0.771
FLMY	0.940
MY305D	0.917
FSP	0.470
DDRY	0.959
FCI	0.997
PY	0.684
DTAP	0.558
MY300D	0.917

[AFC= Age at first calving, LL= lactation length, FLMY= first lactation milk yield, MY305D= milk yield up to 305 days, FSP= first service period, DDRY= date of dry, FCI= first calving interval, PY= peak yield, DTAP= days to attain peak yield and MY300D= milk yield up to 300 days].

Table 4. Estimates of principle components for studied traits using varimax rotation.

	Component			
	1	2	3	4
AFC	-0.182	-0.010	0.657	-0.268
LL	0.417	0.230	0.694	0.250
FLMY	0.852	0.115	0.429	0.130
MY305D	0.957	-0.010	0.047	-0.009
FSP	0.133	-0.015	0.155	-0.654
DDRY	-0.074	0.968	-0.080	-0.104
FCI	0.119	0.963	0.235	0.017
PY	0.784	0.003	-0.261	-0.005
DTAP	0.150	-0.091	0.080	0.722
MY300D	0.956	-0.010	0.048	-0.008

of loading squares was optimized using the varimax rotation technique [20]. The test of Bartlett's sphericity was used to determine the significance of the correlation matrix. Table 1 shows a chi-square value of 7723.339, which is extremely significant ($p<0.01$). PCA was used to extract the sum of squares loadings.

The scree plot provided a visual representation of the eigenvalues associated with the principal components, helping to determine the number of significant components to retain for further analysis. The scree plot showed a steep decline in eigen values from the first to the second component, followed by a gradual flattening of the curve. This pattern suggested that the first few principal components captured most of the variation in the dataset, while later components contributed minimal additional variance. Component 1 has the highest eigenvalue (3.614) and explained 36.141% of the variance, indicating that it was the most influential component. These results suggested that Component 1 captured the overall milk production potential of the animals, with higher values reflecting better milk yield performance over different lactation periods. The highest loadings were observed for MY305D (0.957), MY300D (0.956), FLMY (0.852), and PY (0.784), indicating that these variables strongly contribute to this component. Component 2 accounted for 20.056% of the variance, with a total eigenvalue of 2.006. The component 2 is predominantly associated with reproductive efficiency. The highest loadings are

for DDRY (0.968) and FCI (0.963), indicating that these traits significantly influence this component. Since DDRY represents the drying-off date and FCI reflects the time between successive calvings, Component 2 likely represents reproductive efficiency and calving intervals, with higher values corresponding to longer intervals and delayed drying periods. Component 3 explained 12.620% of the variance, highlighting age at first calving (AFC) and lactation length (LL) as major contributors. These results indicate that Component 3 captures the effect of age and lactation duration on milk yield, with a potential link between early maturity and improved lactation performance. Component 4, contributing 11.113% of the variance,

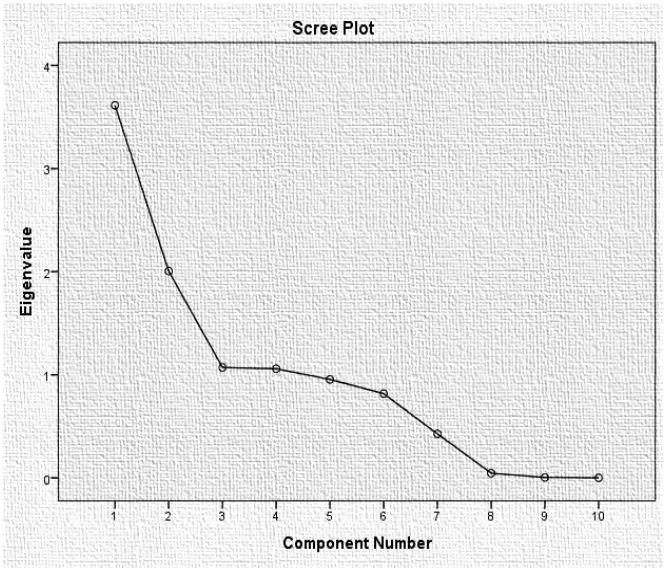

Fig. 1. Scree plot showing component number with eigen value.

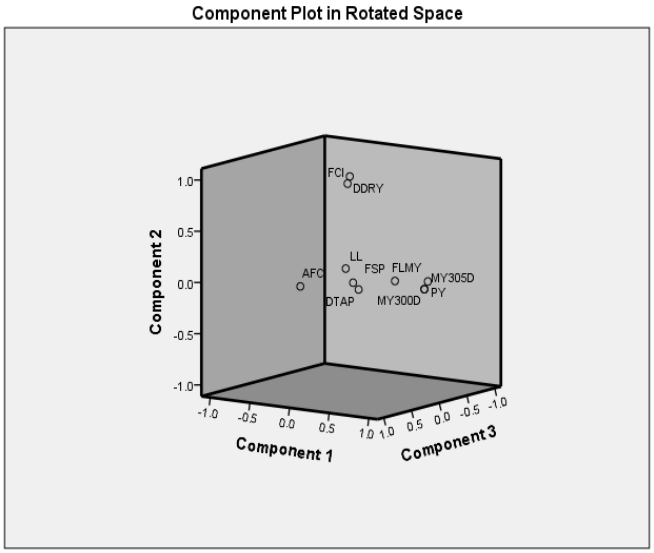

Fig. 2. Component plot in rotated space showing different traits in Crossbred cattle.

was primarily influenced by days to attain peak yield (DTAP) and first service period (FSP). This indicated a connection between the timing of peak production and reproductive performance. The remaining components (PC5-PC10) with the eigen values below 1 contributed minimal variance (<10% each), indicating that their influence was relatively minor and these components were unlikely to contribute significantly to meaningful variation. Given that the first four components together explained a substantial portion of the variance, they provided valuable insights into genetic selection and management strategies for improving milk production and reproductive efficiency in crossbred cattle. [6] also reported that Principal component 1 (PC1) accounted for the majority of overall variance and was heavily influenced by calving interval, service period, and lactation length. Similarly, PC2 demonstrated substantial loading on test day peak yield and milk yield upto 305 days. The third main component PC3 was discovered to have substantial loading on age at sexual maturity and number of services per conception in Frieswal cattle.

The variance of each attribute was as described by PCA [21]. Figure 2 displayed the component plot in rotated space, which illustrated the distribution of the ten components. The component plot in rotated space provided a clearer interpretation of the relationships between traits by redistributing variance more evenly across components. The rotation enhanced interpretability by grouping correlated traits into distinct principal components, helping to identify underlying biological or genetic influences. The traits were positioned within a three-dimensional space formed by component 1, component 2 and component 3. In component 1, traits such as FLMY, MY305D, MY300D and PY were closely grouped, representing milk yield-related parameters, suggesting that this component captured overall production performance. Traits FCI and DDPY were grouped together in component 2. This grouping suggested a relationship between fertility and calving intervals, indicating that this component primarily represented reproductive efficiency. Traits such as AFC, LL, DTAP, and FSP were associated with component 3. These traits were likely related to reproduction and lactation length. The separation of traits into distinct principal components suggested that milk yield and reproductive traits each contributed uniquely to overall genetic variation. The observed clustering indicated that the selection for higher milk yield traits (FLMY, MY305D) might have limited direct correlation with reproductive efficiency (FCI, DDPY), highlighting the need for a balanced selection approach

in breeding programs. The presence of AFC (age at first calving) and LL (lactation length) in a separate component suggested that early reproductive maturity and extended lactation performance were interrelated but distinct from primary milk yield traits.

The estimation of communality for each variable demonstrated the proportion of that variable's variance that could be accounted for by the other components together [17]. The communality values ranged from 0.470 (first service period) to 0.997 (first calving interval) across all production and reproduction traits depicted in table 3. Traits such as first service period, showed lower communality, suggesting that it was not as efficient in elucidating the overall performance in crossbred cows. On the other hand, traits with high communalities, such as lactation milk yield, 305 days milk yield, dry period, calving interval and milk yield upto 300 days for first lactation will be useful in selecting crossbred cattle for breeding programs.

Table 4 illustrated the estimates of principle components using varimax rotation for traits under study. [6] subjected different production and reproduction traits (age at sexual maturity, test day peak yield, milk yield up to 305 days, average percentage of fat, lactation length, calving interval, service period and number of services per conception) of Frieswal cattle to PCA under field progeny testing. Using factor analysis with varimax rotation, three main components were found, and taken together, they accounted for 74.30% of the variation. 35.54% of the variance was explained by PC1, with PC2 and PC3 accounting for 23.37% and 15.38% of the variance, respectively. The communality values for each performance attribute ranged from 0.972 for calving interval to 0.247 for average percentage of fat. [22] conducted a study on performance attributes (calving interval, days open, dry period and lactation length) in Holstein Friesian cows, describing variation as 38.3%, 17.6%, 0.2% and 44.0%, respectively in four principal components. Applying the varimax rotated method, the PC1 coefficients, which encompassed the dry period and lactation time ranged from -0.432 to 0.926. The PC2 coefficients, which vary from -0.072 to 0.892, represented the weights associated with the number of days open and the interval between calving. [5] employed PCA and a correlation matrix to explore the connection between age at first calving, calving interval, reproduction efficiency, overall milk yield, and lactation duration. More than 90% of the variation was explained by the first three principal components, with age at first calving contributing 23.06% and overall milk output accounting for 71.92%. [23] reported their research findings on lifetime performance attributes predicted by PCA, which accounted for 97.244% of the variation. While the PC2 explained 38.948% of the variation, the PC1 explained 58.296%. The communality varied between the peak milk output (0.992) and the overall milk yield (0.955).

For the production parameters of the first main component, the coefficients derived from the varimax rotational technique varied from 0.904 (total milk yield) to 0.436 (peak yield). Principal component regression analysis was employed by [24] to forecast the Jaffarabadi buffaloes' lifetime milk production. They discovered that six early age records *viz*. milk yield, first lactation, total lactation length, peak milk yield, milk yield second lactation. Lactation length and peak milk yield accounted for 98% of the variance in lifetime milk yield. They came to the conclusion that predictions based on these six production potential records of an animal may provide a better foundation for early age selection. Despite numerous advantages of PCA there are certain limitations:- PCA is a dimension reduction approach and it is based on assumptions, therefore may not be generalized in the actual sense and also results in loss of data [25]. The presence of linear correlations (*e.g.*, sphericity test by Bartlett) or sample adequacy (*e.g.*, Kaiser-Meyer-Olkin test) are prerequisites for applying PCA and producing pertinent findings. Since the principal components are intended to be orthogonal to one another, the assumption of orthogonality is another restriction [26].

CONCLUSION

It could be implied that the number of first lactation attributes that needed to be recorded in crossbred cattle were reduced by using orthogonal synthetic variables principal components, specifically PC1, PC2, PC3 and PC4. This information could be utilized to explain production and even reproduction in crossbred cows. The PC1 can be employed in phenotypic selection to explicate the overall milk production potential of the animals, with higher values reflecting better milk yield performance over different lactation periods in crossbred cows to be used in breeding programmes. Future research should focus on validating the Principal Component Analysis (PCA) results with genetic parameters to ensure their biological and genetic relevance. This can be accomplished by calculating heritability estimates to evaluate the genetic contribution of the components, estimating genetic correlations between principal components and the original traits, and performing genome-wide association studies

(GWAS) to find genetic markers associated with the principal components. Additionally, comparisons with traditional selection methods will help evaluate the efficiency of PCA-based selection strategies. Further validation across different populations or breeds is also necessary to confirm the applicability of the identified principal components in genetic improvement programs.

ACKNOWLEDGEMENT

The authors wish to thank the COVAS, Uttarakhand (GBPUA&T, Pantnagar) for collection of data from dairy cattle farm, Nagla, GBPUA&T, Pantnagar.

REFERENCES

1. Tamboli P, Bharadwaj A, Chaurasiya A, Bangar YC, Jerome A. Association between age at first calving, first lactation traits and lifetime productivity in Murrah buffaloes. Anim Biosci. 2022; 35(8), DOI: 10.5713/ab.21.0182.

2. Tamboli P, Bharadwaj A, Chaurasiya A, Bangar YC, Jerome A. Genetic parameters for first lactation and lifetime traits of Nili-Ravi buffaloes. Front Vet Sci. 2021; 8(1), DOI: 10.3389/fvets.2021.557468.

3. Moawed SA, Osman MM. Dimension reduction of phenotypic yield and fertility traits of Holstein-Friesian dairy cattle using principal component analysis. Inter J Vet Sci. 2018; 7(2): 75-81.

4. Abreu BDS, Severino BPB, Elizabete CDS, Kleber RS, Ângela MVB, *et. al.* Principal component and cluster analyses to evaluate production and milk quality traits. Rev Ciênc Agron. 2020; 51(3), DOI: 10.5935/1806-6690.20200060.

5. Mello RRC, Sinedino LDP, Ferreira JE, de Sousa SLG, de Mello MRB. Principal component and cluster analyses of production and fertility traits in Red Sindhi dairy cattle breed in Brazil. Trop Anim Health Prod. 2020; 52(1), DOI: 10.1007/s11250-019-02009-7.

6. Sarma O, Barwal RS, Singh CV, Kumar D, Singh CB *et al*. Principal component analysis in production and reproduction traits of Frieswal cattle under field progeny testing. Pant J Res. 2024; 22(1): 158-163.

7. Moawed SA, Osman MM, Rady EA, El-Bayomi KM, Farag AF. Principle component analysis of breeding values estimated by six animal models for evaluating some productive and reproductive traits of Holstein dairy cattle. Adv Anim Vet Sci. 2021; 9(8), DOI: 10.17582/journal.aavs/2021/9.8.1113.1122.

8. Salem MM, Nasr MA, Amin AM. Principal component analysis of breeding values for birth weight milk and reproductive traits of the Egyptian buffalo. Trop Anim Health Prod. 2021; 53(1), DOI: 10.1007/s11250-021-02625-2.

9. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Phil Trans R Soc A. 2016; 374(20150202), DOI: 10.1098/rsta.2015.0202.

10. Ogwu MC, Osawaru ME. Principal component analysis: A tool for multivariate analysis of genetic variability. African J Plant Sci. 2016; 10(10): 50-52.

11. Amaya A, Martínez R, Cerón-Muñoz M. Selection indexes using principal component analysis for reproductive, beef and milk traits in Simmental cattle. Trop Anim Health Prod. 2021; 53(3), DOI: 10.1007/s11250-021-02815-y.

12. Vargas G, Schenkel FS, Brito LF, de Rezende Neves HH, Munari DP, *et al*. Unravelling biological biotypes for growth, visual score and reproductive traits in Nellore cattle via principal component analysis. Livest. Sci. 2018; 217(1), DOI: 10.1016/j.livsci.2018.09.010.

13. Lever J, Krzywinski M, Altman N. Points of significance: Principal component analysis. Nat Methods. 2017; 14(7): 641-643.

14. Maxwell AF. Statistical methods in factor analysis. Psychol Bull. 1959; 56 (1): 228–235.

15. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis, (Chapter 3). Statistics 459 (6 th. edn.), 1982; USA: Prentice-hall Inc.

16. Everitt BS, Landau S, Leese M. Cluster Analysis, (Chapter 18). Cluster Analysis (4 th edn.), 2001; Arnold Publisher, London.

17. Vohra V, Niranjan SK, Mishra AK, Jamuna V, Chopra A, *et. al*. Phenotypic characterization and multivariate analysis to explain body conformation in lesser known buffalo (*Bubalus bubalis*) from North India. Asian-Australas J Anim Sci. 2015; 28(3), DOI: 10.5713/ajas.14.0451.

18. Sinha R, Verma A, Sinha B, Kumari R, Revanasiddu D, *et al*. Unravelling the relationship between udder morphometric traits and milk production, composition and clinical mastitis in Karan Fries cattle via principal component analysis. Indian J Anim Sci. 2023; 76(3): 268-278.

19. Khargharia G, Kadirvel G, Kumar S, Doley S, Bharti PK, *et. al*. Principal component analysis of morphological traits of Assam hill goat in Eastern Himalayan. Indian J Anim Plant Sci. 2015; 25(5): 1251- 1258.

20. Fernandez G. Data Mining using SAS Application, 2002; USA: Chapman and Hall, CRC press.

21. Mavule BS, Muchenje V, Bezuidenhout, Kunene NW. Morphological structure of Zulu sheep based on principal component analysis of body measurements. Small Rumin Res. 2013; 111(1-3), DOI: 10.1016/j.smallrumres.2012.09.008.

22. Sanad SS, Gharib MG, Ali MAE, Farag AM. Prediction of milk production of Holstein cattle using principal component analysis. J Anim Poult Prod. 2021; 12(1), DOI: 10.21608/jappmu.2021.149198.

23. Ratwan P, Mandal A, Kumar M, Chakravarty AK. Prediction of lifetime performance traits by principal component analysis in Jersey crossbred cattle at an organized farm of eastern India. Indian J Anim Sci. 2017; 87(9): 1163-1167.

24. Dangar N, Vataliya P. Principal component regression analysis to predict lifetime milk yield of Jaffarabadi buffaloes. Buffalo Bull. 2024; 43(3), DOI: 10.56825/bufbu.2024.4334036.

25. Palo HK, Sahoo S, Subudhi AK. Dimensionality reduction techniques: Principles, benefits, and limitations. Data Analytics in Bioinformatics: A Machine Learning Perspective, (Chapter 4). In: Satpathy R, Choudhury T, Satpathy S, Mohanty SN, Zhang X, editors. Data Analytics in Bioinformatics: A Machine Learning Perspectives, 2021; Wiley Online Library. DOI: 10.1002/9781119785620.ch4

26. Libório MP, Da Silva Martinuci O, Machado AMC, Machado-Coelho TM, Laudares S, *at. al.* Principal component analysis applied to multidimensional social indicators longitudinal studies: limitations and possibilities. Geo J 2022; 87(3), DOI: 10.17632/3pp465nfsz.1.

Cite this article as: Kaur S, Ghosh AK, Sarma O, Barwal RS. Application of principal component analysis for evaluating first lactation traits in crossbred cattle. Explor Anim Med Res. 2025; 15(2), DOI: 10.52635/eamr/15.2.241-247.