Print ISSN : 0972-8813
e-ISSN : 2582-2780
[Vol. 22(1) January-April 2024]

Pantnagar Journal of Research

(Formerly International Journal of Basic and Applied Agricultural Research ISSN : 2349-8765)

G.B. Pant University of Agriculture & Technology, Pantnagar

PANTNAGAR JOURNAL OF RESEARCH

Vol. 22(1) January-April 2024

CONTENTS

Principal component analysis in production and reproduction traits of Frieswal cattle under field progeny testing

OLYMPICA SARMA[1], R. S. BARWAL[1*], C. V. SINGH[1], D. KUMAR[1], C. B. SINGH[1], A. K. GHOSH[1], B. N. SHAHI[1] and S. K. SINGH[2]

[1]*Department of Animal Genetics and Breeding,* [2]*Department of Livestock Production and Management, College of Veterinary and Animal Sciences, G. B. Pant University of Agriculture and Technology, Pantnagar-263145 (U. S. Nagar, Uttarakhand)*

[*]*Corresponding author's email id: drbarwalrs63@gmail.com*

ABSTRACT: Principal Component Analysis is a mathematical procedure employed to transform a set of correlated variables into a smaller set of uncorrelated variables, thereby reducing the dimensionality. By applying principal component analysis, a comprehensive data set comprising various production and reproduction parameters such as milk yield, age at first calving, and calving interval etc. can be effectively analyzed. The study aimed to analyze the principal component of production and reproduction traits in Frieswal cattle, data spanning from 2013 to 2021, comprising of production and reproduction traits of 1163 cattle across six different field units, were collected and subjected to PCA to explain the performance in Frieswal. Factor analysis with varimax rotation uncovered three principal components, collectively explaining 74.30% of the total variance. The first principal component accounted for 35.54% of the variance, followed by the second and third components, which explained 23.37% and 15.38% of the variance, respectively. The communality values ranged from 0.247 (average fat %) to 0.972 (calving interval) across all performance traits. These findings indicate that PCA can serve as a valuable tool in breeding programmes, allowing for a significant reduction in the number of production and reproduction traits while still effectively capturing performance trends in Frieswal cattle.

Key words: Frieswal cattle, PCA, production traits, reproduction traits

The success of indigenous dairy cow breeds depends on the animals' production and reproduction performance levels (Patel *et al.*, 2022). Production traits encompass a range of characteristics related to the quantity and quality of milk produced, as well as the reproductive traits comprised of age at first calving, age at sexual maturity, calving interval, service period, number of services per conception etc. and among these, age at first calving, calving interval and service period are considered as most important traits. Production and reproduction traits are used to characterize different breeds of livestock as they give an idea of performance in animals. These traits are essential for optimizing farm productivity, meeting market demands, and ensuring the economic viability of dairy enterprises.

The rapid advancements in research have led to a significant increase in the volume of large data sets. Interpreting such extensive data within a short time frame with minimal error has become a considerable challenge. Therefore, to overcome this problem a large set of techniques have been developed which leads to fulfill this requirement; however, principal component analysis (PCA) is most commonly used technique (Jolliffe and Cadima, 2016). PCA is a mathematical procedure that transforms a number of possibly correlated variables into smaller number of uncorrelated variables leading to dimension reduction. It is a multivariate statistical tool that examines a data table where observations are characterized by several inter-correlated quantitative dependent variables (Abdi and Williams, 2010). It can be employed for simplifying data, reducing data sets, classifying data, selecting variables, and various other purposes (Wold *et al.,* 1987). The primary purpose of PCA is to extract essential information from the data and represent it through a set of summary indices (PC1, PC2, PC3 etc.) referred to as principal components. The feature representing the direction with the highest variance in the data is assigned as the first Principal Component (PC1). The trait indicating the second most significant source of variation in the data is identified as the second

Principal Component (PC2). Together, these two principal components form a plane that can be visually represented graphically.

Frieswal is a recognized synthetic breed originated from a strategic crossbreeding having 5/8 Holstein Friesian and 3/8 Sahiwal inheritance. This breed is available in Uttar Pradesh and Uttarakhand states of India and the outstanding performance of Frieswal cows in diverse locations demonstrates the adaptability of this germplasm to a range of agro-climatic conditions in India. Apart from their outstanding milk production, Frieswal cattle showcase favourable reproductive characteristics, showcasing a calving efficiency with approximately 80% of these animals calving before reaching three years of age. This combination of excellent production and reproductive qualities, positions the Frieswal as a valuable asset in the realm of dairy farming (Annual Report, ICAR-CIRC, 2019). The current research was conducted to examine the productive and reproductive traits of Frieswal cattle and to develop components for determining which measures most accurately represent the performance in Frieswal cattle.

MATERIALS AND METHODS

The current research is conducted on ongoing Field Progeny Testing (FPT) programme involving Frieswal cattle, a synthetic breed, in Udham Singh Nagar district of Uttarakhand. Data pertaining to production and reproduction traits were extracted from records. The progeny testing programme commenced at the Pantnagar centre in 2009, serving as one of the facilities under the pre-existing All India Coordinated Research Project (AICRP) on progeny testing initiated by ICAR-CIRC, Meerut (Uttar Pradesh). Data consisted of eight different performance traits (production and reproduction) on 1163 Frieswal cattle from six different field units. The data utilized in the current study were derived from records documenting milk production and reproduction traits observed in the progeny of the tested sires over the period from 2013 to 2021. The recorded measurements included in the study were lactation length (LL), 305- day milk yield (305D-MY), average fat percentage (FP), test day peak yield (TDPY), age at sexual maturity (ASM), number of services per conception (NSPC), calving interval (CI) and service period (SP).

Statistical Analysis

The phenotypic correlations between production and reproduction traits were estimated by using Pearson correlation (Snedecor and Cochran, 1967). Traits showing high correlation were then subjected to multivariate principal component analysis. The goal of principal component analysis is to capture the highest proportion of the variance found in the initial set of variables using a minimal number of composite variables. This approach assumes that the communalities (common variance) constitute a minor portion of the overall variance. Varimax rotation was applied to the principal components to transform them into a simpler structure. Initially, Bartlett's test (1950) was applied to determine whether the dataset comprising 1163 animals and eight traits could be subjected to factorization, as recommended by Maxwell (1959). At a significance level of 1%, the data set's validity was confirmed through the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy. The criterion set by the Kaiser rule (Johnson and Wichern, 1982) was employed to ascertain the number of factors, retaining only those with eigen values more than 1. Kaiser's measure of sampling adequacy (MSA) was utilized to assess the suitability of the common factor model, with an MSA below 0.5 considered unacceptable. Principal component analysis, as outlined by (Everitt *et al.,* 2001), is a technique used to transform variables in a multivariate dataset (X_1, X_2, …, X_n) into a set of uncorrelated variables (Y_1, Y_2, …, Y_n). These new variables explain a reduced proportion of the total variance present in the original variables specified as:

$$Y_1 = a_{11}X_1+a_{12}X_2+\ldots+ a_{1n}X_n.$$
$$Y_2 = a_{21}X_1+a_{22}X_2+\ldots+ a_{2n}X_n.$$
$$Y_n = an_1X_1+an_2X_2+\ldots+ a_nX_n$$

The principal components Y_1, Y_2, ..., Y_n represents reduced fractions of the overall variance observed in the original variables X_1, X_2, ..., X_n. Orthogonal rotation, aimed at maximizing variance, was employed in the linear transformation of the factor pattern matrix. This was done to facilitate a clearer interpretation of the extracted principal components. The analyses for principal components were conducted using the factor module within the SPSS 24.

RESULTS AND DISCUSSION

Principal component analysis was applied to eight different production and reproduction traits in Frieswal cattle which generated eight principal components. The Kaiser-Meyer-Olkin method yielded a measurement of sampling adequacy of 0.697 (Table 1) whereby eigen values more than 1 were considered. This value assesses the adequacy of the data for each factor in providing reliable results for PCA. The measure of sampling adequacy below 0.5 is considered to be inadequate (Khargharia *et al.,* 2015). The varimax rotation method was employed to maximize the sum of loading squares (Fernandez, 2002). Bartlett's test of sphericity was applied to assess the significance of the correlation matrix. The resulting chi-square value was highly significant ($P<0.01$), measuring 6231.055 (Table 1). The sum of squares loadings were extracted by PCA, variation explained by each component (Table 2) and Eigen values are given in Fig 1. Three, out of the eight components were selected using the Kaiser Rule Criterion (Johnson and Wichern, 1982) to determine the significant number of components. The cumulative variance of 74.30 % (Table 2) was accounted for by three principal components (PC1, PC2, and PC3), each with eigen values more than 1. The variance of each trait was as per explained by PCA (Mavule *et al.,* 2013). The component plot in rotated space depicts the distribution of the eight components, as illustrated in Figure 2.

In the present study, the first principal component (PC1) explained for 35.54% of the total variance (Table 2). First principal component (PC1) described the calving interval, lactation length, service period, test day peak yield, 305 days milk yield, average fat % in Frieswal cattle. It was represented by a very high component loading for calving interval, service period and lactation length. The second principal component (PC2) explained for 23.37% of total variance. Second principal component (PC2) described the calving interval, lactation length, service period, test day peak yield, 305 days milk yield, average fat %, number of services per conception. PC2 showed high loadings on test day peak yield and 305 days milk yield. The third principal component (PC3) explained 15.38% of total variance explained loadings on average fat %, age at sexual maturity and number of services per conception which described high loads on age at sexual maturity and number of services per conception. Table 4 represents the coefficient of principal component analysis of rotated component matrix. The different weights were assigned by PC1, PC2 and PC3 showing different component loadings in Frieswal Cattle. Table 3 represents the communality values ranged from 0.247 (average fat %) to 0.972 (calving interval) across all performance traits. The trait like average fat % had lower communality indicating that this trait is less effective in explaining the performance in Frieswal cattle and traits including test day peak yield, 305 days milk yield, lactation length, calving interval and service period showed high communalities indicating that these traits will be effective in breeding programme for selection of Frieswal cattle.

The findings of Sanad *et al.* (2021) described the study including performance traits (Lactation length, calving interval, days open and dry period) in Holstein Friesian cattle explaining the variation for four principal components as 44.0%, 38.3%, 17.6% and 0.2%, respectively. By applying varimax rotated method, the coefficients of the first principal

Table 1. KMO and Bartlett's Test

Kaiser-Meyer-Olkin	Measure of Sampling Adequacy	0.697
	Approx. Chi-Square	6231.055
Bartlett's Test of Sphericity	Df	28
	Sig.	0.000

Table 2: Total variance explained by different factor

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.113	38.911	38.911	3.113	38.911	38.911	2.843	35.544	35.544
2	1.624	20.294	59.205	1.624	20.294	59.205	1.870	23.378	58.921
3	1.208	15.099	74.304	1.208	15.099	74.304	1.231	15.383	74.304
4	.892	11.145	85.449						
5	.770	9.623	95.072						
6	.239	2.987	98.059						
7	.115	1.441	99.500						
8	.040	.500	100.000						

component, ranging from -0.432 to 0.926, encompassing dry period and lactation length. In contrast, the coefficients of the second principal component, ranging from -0.072 to 0.892, signify the weights associated with days open and calving interval. Mello *et al.* (2019) applied PCA using a correlation matrix to explore the relationship between age at first calving, calving interval, reproductive efficiency, total milk yield and lactation length explaining a total variation of more than 90% by the first three principal components among which total milk yield accounted for 71.92%, while age at first calving contributed 23.06% of the total variation. Ratwan *et al.* (2017) revealed its findings by conducting a study on lifetime performance traits predicted by PCA which explained a total variation of 97.244%. The first principal component explained a variation of 58.296% and the second principal component showed a variation of 38.948%. The

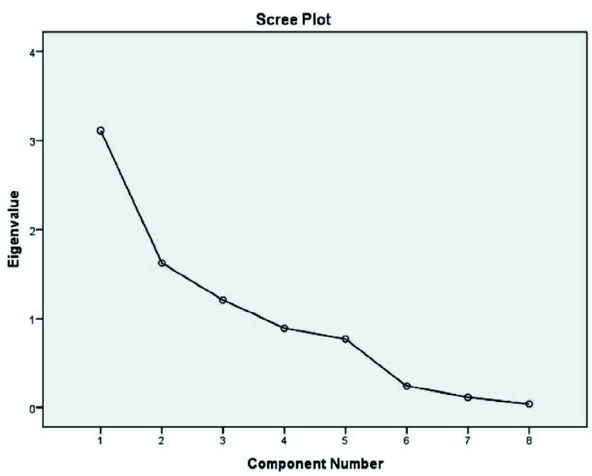

Fig 1: Scree plot showing component numbers with eigen values

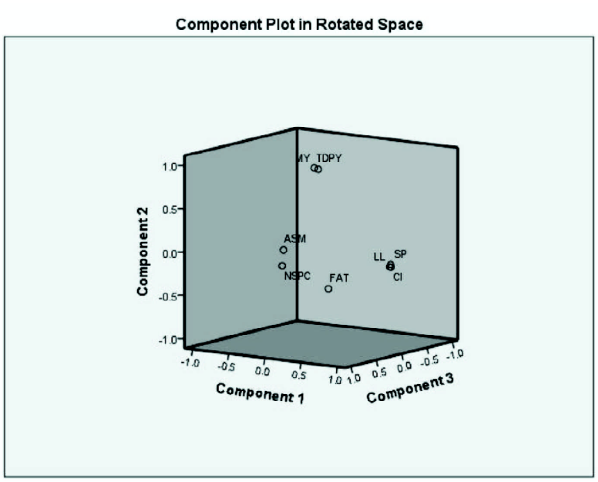

Fig 2: Component plot in rotated space showing six different components.

Table 3: Communalities of different traits

Traits	Extraction
ASM	0.631
TDPY	0.802
305D-MY	0.834
FP	0.247
LL	0.939
CI	0.972
SP	0.930
NSPC	0.589

ASM: Age at sexual maturity, TDPY: Test Day peak yield, 305D-MY: 305 days milk yield, FP: average fat percentage, LL: Lactation length, CI: Calving interval, SP: Service period, NSPC: Number of services per conception

Table 4: Varimax rotated component matrix showing different component loadings in Frieswal Cattle

Component	1	2	3
1	0.905	-0.420	0.069
2	0.425	0.887	-0.181
3	0.015	0.193	0.981

communality ranged between 0.955 (Total milk yield) and 0.992 (peak yield). The coefficients obtained by varimax rotated method for the production traits of first principal component ranging from 0.904 (total milk yield) to 0.436 (peak yield).

CONCLUSION

The study suggests that principal components provide an effective approach for selecting animals by considering a set of inter-correlated variables. The three extracted factors determine the underlying source of shared variability in explaining the performance in Frieswal cattle. Principal component 1 (PC1) had the largest share of overall variance and had high loading on calving interval, service period and lactation length. Similarly, principal component 2 (PC2) seemed to have high loads on test day peak yield and 305 days milk yield. The subsequent principal component 3 (PC3) was found to have high loads on age at sexual maturity and number of services per conception. The three main components discovered in the study could be used as a factor score to predict various production and reproduction traits. Therefore, the study revealed that the components extracted could be used in breeding programmes with reduction in the number of production and reproduction traits to be recorded for explanation of maximum variability for prediction of performance in Frieswal cattle.

REFERENCES

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4): 433-459.

Annual Report, (2019). ICAR-Central Institute for Research on Cattle, Meerut Cantt (Uttar Pradesh)

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, 3: 77–85.

Everitt, B. S., Landau, S. and Leese, M. (2001). Cluster Analysis. 4th edn. Arnold Publisher, London.

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

Johnson, R. A. and Wichern, D.W. (1982). Applied Multivariate Statistical Analysis. USA: Prentice-hall Inc.

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences,* 374 (2065): 20150202.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23: 187–200.

Khargharia, G., Kadirvel, G., Kumar, S., Doley, S., Bharti, P. K. and Das, Mukut. (2015). Principal Component Analysis of Morphological Traits of Assam Hill Goat in Eastern Himalayan India. *Journal of Animal & Plant Science,* 25 (5): 1251- 1258.

Mello, R., Sinedino, L. D., Ferreira, J. E., de Sousa, S. and de Mello, M. (2019). Principal component and cluster analyses of production and fertility traits in Red Sindhi dairy cattle breed in Brazil. *Tropical Animal Health and Production*, 52(1): 273–281.

Mavule, B.S., Muchenje, V., Bezuidenhout and Kunene, N.W. (2013). Morphological structure of Zulu sheep based on principal component analysis of body measurements. *Small Ruminant Research,* 111: 23-30.

Maxwell, A. F. (1959). Statistical methods in factor analysis. *Psychological Bulletin*, 56 (1): 228–235.

Patel, A. K., Kachhawah, S. C., Patil, N. V. and Chopra, A. (2022). Effects of Non-genetic Factors on Production and Reproduction Traits in Tharparkar Cattle under Arid Region of Rajasthan. *Indian Journal of Animal Research,* 56(3): 263-269.

Ratwan, P., Mandal, A., Kumar, M. and Chakravarty, A. K. (2017). Prediction of lifetime performance traits by principal component analysis in Jersey crossbred cattle at an organized farm of eastern India. *The Indian Journal of Animal Sciences*, 87(9): 1163-1167.

Sanad, S. S., Gharib, M. G., Ali, M. A. E. and Farag, A. M. (2021). Prediction of Milk Production of Holstein Cattle Using Principal Component Analysis. *Journal of Animal and Poultry Production*, 12(1): 1-5.

Snedecor, G.W. and Cochran, W.G. (1967). Statistical Methods. Oxford and IBH Publishing Co., New Delhi, India.

SPSS. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Wold, S., Esbensen, K. and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2 (1-3): 37-52.

Received: March 18, 2024
Accepted: April 10, 2024

.