RESEARCH ARTICLE Indian Journal of Animal Research, Volume 58 Issue 8: 1418-1422 (August 2024)

The Evaluation of Relationships between Milk Composition Traits and Breeds with Categorical Principal Component Analysis in Akkaraman and Awasi Sheep

Bahattin Çak[1], Sıddık Keskin[2], Gökhan Aydemir[3]

10.18805/IJAR.BF-1791

ABSTRACT

Background: This study aims to determine the relationship between milk composition traits and breed in the Akkaraman and Awasi sheep as well as to provide ease of interpretation by showing the relationships structure between variables and between categories of variables in two-dimensional space with Categorical principal component analysis.

Methods: Categorical principal component analysis determines relationships between continuous and categorical variables as well as ordinal variables. It aims to reduce system dimensionality through optimal scaling while maintaining variable measurement levels (nominal, multiple nominal, ordinal and interval). In this research, data obtained from Akkaraman and Awasi Breed Sheep Raised by Public Hands in Tuşba District of Van Province were used. In order to determine relationship with breed, the traits were divided into two categories, “low” and “high” and all variables (9 variables) were considered together and a Categorical principal components analysis was performed.

Result: As a results, Dimension 1 accounted for 35.58% of the total variation while dimension 2 accounted for 15.21%. Two dimensions together accounted for 50.79% of the variation. Thus it can be noted that Categorical principal component analysis can be used in the analysis of data sets containing a large number of different types of variables with linear or non-linear relationships between them.

Key words: Animal husbandry, Configuration, Dimension reduction, Milk components.

INTRODUCTION

According to the estimates of FAO, world total milk production in 2021 amounted to 928 million tons, an increase of 1.5% compared to the previous year (Ulusal Süt Konseyi, 2021). In 2022, 83.3% of the animals slaughtered, 27.6% of the meat produced and 7.5% of the milk produced were obtained from small ruminants (TUIK, 2022). There are 1.28 billion sheep and 1.11 billion goats in the world (FAOSTAT, 2021). 3.48% of the world’s sheep and 1.04% of the world’s goats are in Türkiye. Türkiye ranks 6th in the world sheep population and 20th in the goat population (FAOSTAT, 2021). CATPCA (Levy, 2003; Linting, *et al.*, 2007; Horan, *et al.*, 2008; Linting and van der Kooij, 2012), is a method that minimizes the amount of data lost when reducing a set of numerical variables to a smaller number of uncorrelated components. It is derived from linear principal component analysis (PCA). Nominal and ordinal variables can all be converted into quantitative variables using CATPCA’s “optimal quantification” procedure. Although two or three components that contribute a larger amount of variation are typically employed, CATPCA generates as many components as variables included in the model.

Because of its unique nutritional richness and significance for both human and animal health, milk is an essential food. It possesses every element required by living things in the most easily attainable form (Popescu and Angel 2009). Therefore, milk composition traits such as fat, density, lactose and pH are important for the quality of milk. These traits are likely to be related with breeds.

In addition to the relationships between milk composition traits, correct determination of the relationship structure between these traits and breeds is very important for selection studies to be carried out to obtain the desired breeds in terms of these traits. Therefore, this study aims to examine the relationship structure between milk components as well as the relationship structure between these components and breeds.

[1]Department of Animal Science, Faculty of Veterinary, Van Yüzüncü Yıl University, 6508, Van, Türkiye.

[2]Department of Biostatistics, Faculty of Medicine, Van Yüzüncü Yıl University, 6508, Van, Türkiye.

[3]T.C. Ministry of Agriculture and Forestry, Ceylanpınar District Directorate of Agriculture, Şanlıurfa, Türkiye.

Corresponding Author: Bahattin CAK, Department of Animal Science, Faculty of Veterinary, Van Yüzüncü Yıl University, 6508, Van, Türkiye. Email: bahattincak@yyu.edu.tr

How to cite this article: Çak, B., Keskin, S., Aydemir, G., (2024). The Evaluation of Relationships between Milk Composition Traits and Breeds with Categorical Principal Component Analysis in Akkaraman and Awasi Sheep. Indian Journal of Animal Research. 58(8): 1418-1422. doi: 10.18805/IJAR.BF-1791.

Submitted: 27-03-2024 **Accepted:** 04-05-2024 **Online:** 12-06-2024

MATERIALS AND METHODS

The study was carried out on Akkaraman (60 head) and Awassi breed (60 head) sheep in two different private sheep breeding enterprises in Tuşba district of Van province. The chemical (fat, protein, non-fat solids and lactose) and physical (specific gravity, freezing point, electrical conductivity and pH) properties considered in the study were obtained from milk samples drawn from 1st, 2nd, 3rd and 4th month of lactation periods. Analyzes for all traits were performed at Van Yüzüncü Yıl University, Faculty of Veterinary Medicine, Laboratory of Animal Science. Then, the mean of these values was considered to the determination of the relationships. Thus, the possible differences in the milk composition traits in different periods of lactation was tried to be eliminated. In order to determine relationship with breed, the traits were divided into two categories, "low" and "high" and all variables (9 variables) were considered together and a Categorical principal components analysis was performed.

Statistical analysis

Categorical principal component analysis is a dimension reduction method. It provides the opportunity to visually present linear and non-linear relationships both between variables and between categories of variables in a two-dimensional space (Abou-Senna *et al.,* 2021).

n: number of observations or subjects

m: number of variables (features).

Let H: be an *nxm* dimensional data matrix. The variable in each column of the matrix H can be represented by the *nx1* dimensional vector hj (*j* = 1, 2,…,m). If any variable hj is not a continuous variable; Most likely, the relationship between them is also non-linear. Therefore, it is necessary to apply a non-linear transformation. The transformation of each category includes an optimally scaled value. Thus, the matrix H is replaced by the matrix Q*ij*. The Q matrix contains scaled values for categorical variables instead of their observed scores (Torres-Cárdenas *et al.,* 2021).

Categorical principal component analysis is the same as standard or classical principal component analysis (PCA). It aims to find the relationship structure between the transformed variables. For this purpose, the loss function is tried to be minimized in Categorical principal component analysis. The loss function contains values weighted according to multiple nominal transformations. The score values of objects or individuals in the obtained principal components are also called object scores or score values in the categorical principal components analysis. These components, multiplied by the optimal set of weights, are determined as component saturation and an attempt is made to converge as much as possible to the original data set (Torres-Cárdenas *et al.,* 2021).

Let p is the number of components and X is nxp dimensional component scores matrix and A is mxp dimensional component saturations matrix.

When *j*. row of this matrix is expressed by a*j*, the difference between the original data and the principal components in the Loss function is minimized. Thus the loss function can be written as follows ((Torres-Cárdenas *et al.,* 2021);

$$L(Q,A,X) = n^{-1} \sum_{j=1}^{m} \text{tr}\,(q_j a'_j - x),(q_j a'_j - x)$$

This loss function is subject to the following restrictions: Transformed variables are standardized. Thus $q'_j q_j = n$. This restriction is in the product $q_j\, a'_j$; it is necessary to resolve the uncertainty between q_j and a'_j.

With this normalization; q_j contains the z scores and thus provides the correlation between the component saturations in a_j and the variables as well as components.

To avoid trivial solution; the object scores need to be centralized with the conditions A = 0, X = 0 and X′X = nI (I, identity matrix). Thus, 1′X = 0 (1, unit vector) (Torres-Cárdenas *et al.,* 2021). SPSS (ver: 22) statistical package program was used for statistical analysis.

RESULTS AND DISCUSSION

The Pearson correlations between milk components are given separately in Akkaraman and Awassi breeds in Table 1. As seen in Table 1, the highest correlation is observed between Fat free dry matter and lactose. This is followed by negative correlations between fat free dry matter and Freezing point as well as fat free dry matter and protein in the Akkaraman breed.

Table 1: Correlations between the traits considered in the study for Akkaraman and Awasi breeds.

Akkaraman/Awasi	Fat	Fat freedry matter	Density	Protein	Lactose	Freezingpoint	Conductivity	pH
Fat	1	0.117	-0.145	-0.125	-0.012	-0.291*	-0.151	0.090
Fat free dry matter	0.070	1	0.248	0.341**	0.538**	-0.372**	-0.152	0.125
Density	0.127	0.342**	1	0.237	0.292*	-0.156	0.090	0.100
Protein	-0.032	0.088	0.052	1	0.271*	-0.122	0.153	0.135
Lactose	0.123	0.195	0.428**	0.224	1	-0.171	0.009	-0.028
Freezing point	-0.065	-0.150	0.053	0.086	0.077	1	0.104	-0.103
Conductivity	-0.095	0.075	0.032	0.037	0.087	0.181	1	0.037
pH	0.128	-0.292*	-0.261*	-0.088	0.001	0.006	0.036	1

*: p<0.05; **: p<0.01.

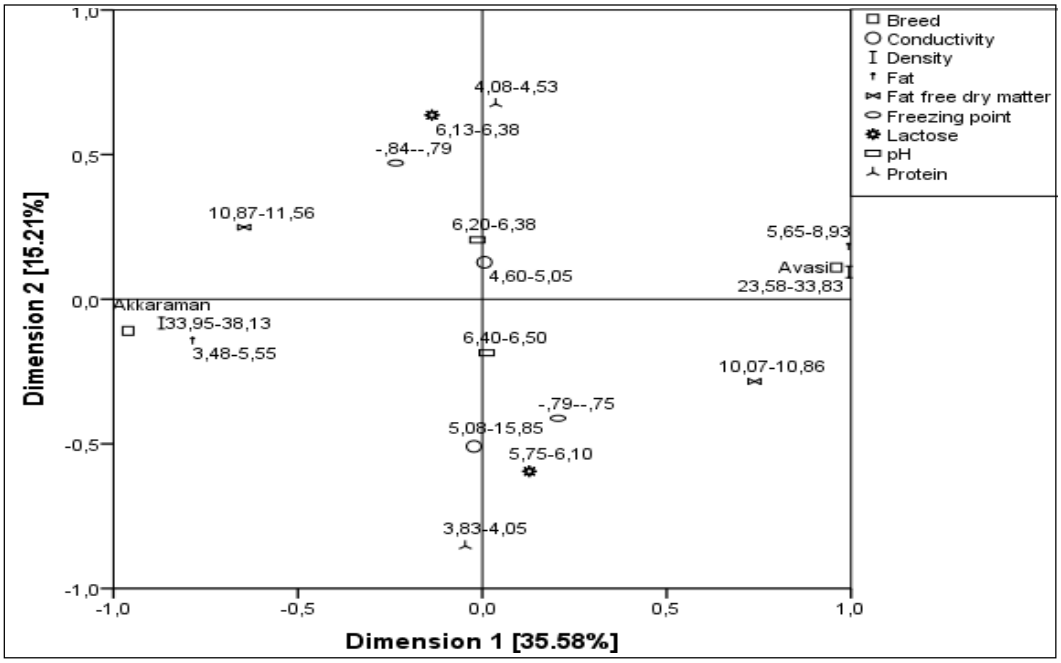


Fig 1: Configuration of the relationship structure between categories of variables in 2 dimensional space.

On the other hand, in the Awassi breed, the highest correlation was observed between lactose and density (0.428), followed by a correlation of 0.342 between density and fat free dry matter. There are low and negative correlations were observed between pH and Fat free dry matter as well as density.

As a result of the Categorical principal component analysis, the configuration of the relationship structure in two-dimensional space for the categories of variables is presented Fig 1. In addition, the configuration of the relationship between the variables is shown in Fig 2. For nominal, ordinal and numerical variables, the modified variables are displayed as vectors in the resulting graph. The correlation between the component and the variable is expressed by the vector length or component loadings, which also serve as an indicator of variance accounted for (VAF) and the component's contribution. A high correlation between variables is shown by an angle near to zero, no association is indicated by a 90° angle and an inverse relationship is indicated by a 180° angle. The correlation coefficient between the vectors is represented by the cosines of the angles between them. Centroids are created as the graphic representation for categorical variables, one for each category. The relationship between the categories is indicated by the centroids' proximity to one another. The vectors and the locations of the centroids of the category variables can be combined to create a graph. (Carreño Renes *et al.,* 2022).

As seen in Fig 1, Dimension 1 accounted for 35.58% of the total variation while dimension 2 accounted for 15.21%. Two dimensions together accounted for 50.79% of the variation. Similarly, Sankhyan *et al.* (2017), conducted a principal component analysis to examine the relationship between 12 traits in 728 sheep and as a result of the analysis, they emphasized that 3 and 4 principal components (factors) accounted for 57% and 61% of the total variation, respectively. Likewise, Mishra *et al.,* (2022), conducted a principal component analysis on 9 traits in Chitarangi sheep and as a result of the analysis, they stated that 3 principal components accounted for 69.06% of the total variation and emphasized that principal component analysis could be used for dimension reduction for biometric traits in sheep.

Categories located on the right and left sides of the graph, being highly negatively correlated with respect to the first dimension. According to first dimension, which accounts for 35.58% of the variance, Awasi is located in the positive region, while Akkaraman is in the negative region. According to the first dimension, the high categories of freezing point and fat as well as the low categories of fat-free dry matter, lactose and density were located in the same region as the Awasi breed. Milk composition of the Awasi breed is positively associated with high categories of freezing point and fat. Similarly milk composition of the Awasi breed is highly and positively correlated with low categories of density, lactose and fat-free dry matter. Thus, it can be stated that the milk composition of the Awasi breed tends to be low density, lactose and fat-free dry matter while high fat and freezing point.

It was observed that lactose, density and fat-free dry matter tend to be high, while fat and freezing point values tend to be low in milk components of Akkaraman breed. pH and protein are highly correlated with the second dimensions. The high category of pH and the low category

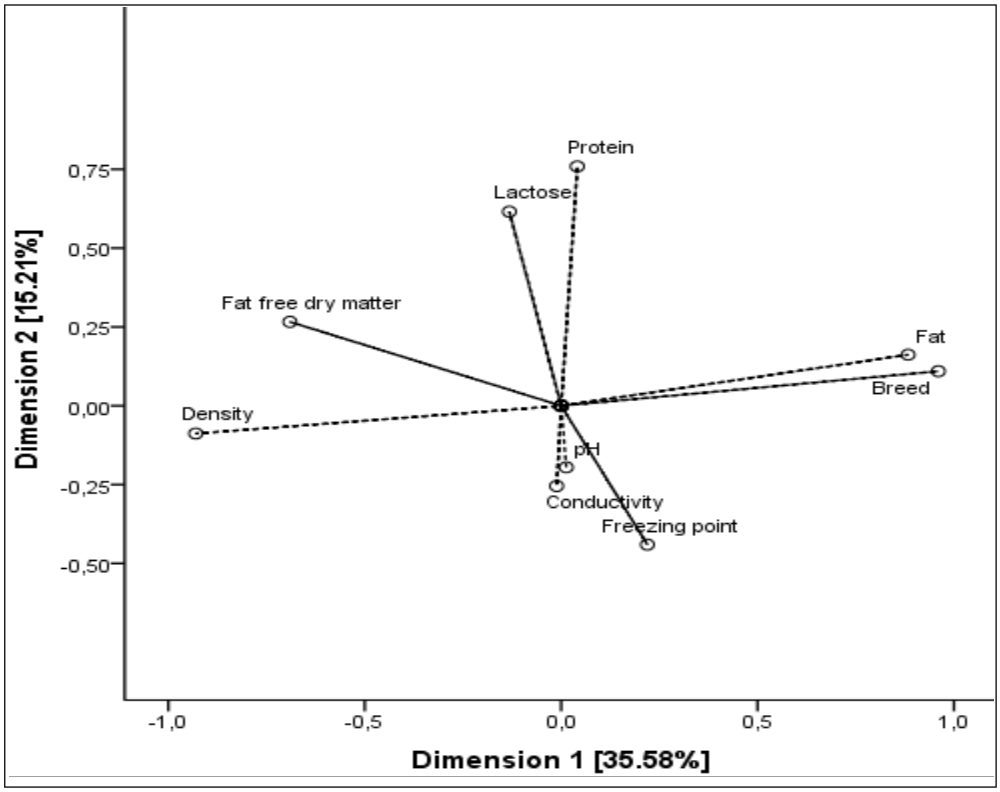


Fig 2: Configuration of the relationship structure between variables in 2 dimensional space.

of protein are located in the negative region of the second dimensions, while the high categories in the positive region.

The configuration of the relationships between variables in two-dimensional space is given in Fig 2. As seen in Fig 2, according to the first dimension, density is highly and positively correlated with fat-free dry matter. Similarly, according to the first dimension, these two variables are highly and negatively correlated with fat and breed. According to the second dimension, a highly and positive relation is observed between protein and lactose while a moderate positive correlation was observed between pH, conductivity and freezing point.

Protein and lactose is negatively correlated with and pH, conductivity and freezing point. Thus it can be stated that the protein and lactose levels of milk components increase while pH, conductivity and freezing point tend to decrease. Similarly, while fat-free dry matter and freezing point increase, the fat level tends to decrease.

All animals' milk contains lipids; however, the amount varies greatly between species, ranging from less than 2% to more than 50%. The primary purpose of dietary lipids for neonates is to provide energy and the amount of fat in milk primarily reflects the energy needs of the species; for example, cold-adapted land animals and marine mammals secrete large amounts of lipids in their milk (Fox *et al.,* 2015b). With two notable exceptions-the California sea lion and the hooded seal-lactose is the main carbohydrate found in the milk of most mammals. Other sugars found in milk in trace levels include fructose (50 mg/l), glucose (50 mg/l), galactosamine, glucosamine and N-acetyl neuraminic acid, which are components of glycolipids and glycoproteins. All examined species' milk contains oligosaccharides, which are important components of some species' milk, such as human milk. (Fox *et al.,* 2015c). Protein content in normal bovine milk is roughly 3.5%. The whey protein fraction experiences the most concentration shift during lactation, which happens within the first few days after delivery. Milk proteins naturally provide young mammals with a variety of biologically active proteins, such as immunoglobulins, zinc- and vitamin-binding proteins and different protein hormones, as well as the essential amino acids needed for the development of muscular and other protein-containing tissues. Diverse species' young have diverse physiological and nutritional needs because they are born at extremely different stages of maturation. The protein content of the species' milk varies, ranging from approximately 1 to 24%, which reflects these variances. Since the young of that species need protein to grow, there is a direct correlation between the protein content of milk and the rate of growth of the young. (Fox *et al.,* 2015d). Aqueous colloidal continuous phase and oil/fat dispersion phase make up the diluted emulsion known as milk. The physical traits of milk are comparable to those of water, however they are altered by the degree of dispersion of the colloidal and emulsified components as well as the presence of different solutes (proteins, lactose and salts) in

the continuous phase. The primary physical traits of milk include its density, conductivity, thermal traits, rheological behavior, redox traits, colligative traits, surface activity buffering capacity and color (Fox *et al.,* 2015a).

CONCLUSION

According to the first dimension, the milk composition of the Awasi breed was observed to be positively correlated with high freezing point and fat categories, while it was also highly and positively correlated with low density categories, lactose and non-fat dry matter. Thus, it can be stated that the milk composition of the Awasi breed tends to be low density, lactose and fat-free dry matter while high fat and freezing point. It was observed that lactose, density and fat-free dry matter tend to be high, while fat and freezing point values tend to be low in milk components of Akkaraman breed. pH and protein were found highly correlated with the second dimension. It has been determined that the high pH category and low protein category are located in the negative region of the second dimension, while the high categories are located in the positive region. As a result, in the study, the relationships between milk composition traits of two breeds were examined by Categorical principal component analysis. The relationship structure between 9 variables has been reduced to 2 dimensions with a variance explanation rate of approximately 51% and is presented in a visually more easily understandable and interpretable way. Thus, it can be concluded that Categorical principal component analysis can be used in the analysis of data sets containing a large number of different types of variables with linear or non-linear relationships between them.

Conflict of interest

The authors have no conflict of interests related to this publication.

REFERENCES

Abou-Senna, H., Radwan, E., T. Abdelwahab, H.T. (2021) Categorical principal component analysis (CATPCA) of pedestrian crashes in Central Florida, Journal of Transportation Safety and Security. doi: 10.10801 9439962.2021.1988788.

FAOSTAT, Food and Agriculture Data. (2021). Crops and Livestock Products. https://www.fao.org/faostat/en/ #data/ QCL. Access: Mayıs, 2023.

Fox, P.F., Uniacke-Lowe, T., McSweeney, P.L.H., O'Mahony, J.A. (2015a). Physical Properties of Milk. In: Dairy Chemistry and Biochemistry. Springer, Cham. https://doi.org/10.10 07/978-3-319-14892-2_8.

Fox, P.F., Uniacke-Lowe, T., McSweeney, P.L.H., O'Mahony, J.A. (2015b). Milk Lipids. In: Dairy Chemistry and Biochemistry. Springer, Cham. https://doi.org/10.1007/978-3-319-14892-2_3.

Fox, P.F., Uniacke-Lowe, T., McSweeney, P.L.H., O'Mahony, J.A. (2015c). Lactose. In: Dairy Chemistry and Biochemistry. Springer, Cham. https://doi.org/10.1007/978-3-319-14892-2_2.

Fox, P.F., Uniacke-Lowe, T., McSweeney, P.L.H., O'Mahony, J.A. (2015d). Milk Proteins. In: Dairy Chemistry and Biochemistry. Springer, Cham. https://doi.org/10.1007/978-3-319-14892-2_4.

Horan, T.C. andrus, M. and Dudeck, M.A. (2008). CDC/NHSN surveillance definition of health care- associated infection and criteria for specific types of infections in the acute care setting. Am. J. Infect. Control. 36: 309-332.

Levy M.M., (2003). 2001 sccm/esicm/accp/ats/sis international sepsis definitions conference. Crit. Care Med. 31(4): 1250-1256.

Linting, M., Meulman, J.J., Groenen, P.J. and van der Koojj, A.J. (2007). Nonlinear principal components analysis: Introduction and application. Psychological Methods. 12(3): 336-358.

Linting, M. and van der Kooij, A. (2012). Nonlinear principal components analysis with CATPCA: A tutorial. Journal of personality assessment. 94(1): 12-25.

Mishra A.K., Jain Anand, Singh S., Pundir R.K. (2022). Study of body conformation of carpet wool type chitarangi sheep of India using principal component analysis. Indian Journal of Animal Research. 56(3): 375-379. doi: 10.18805/IJAR.B-4285.

Popescu, A. and Angel, E. (2009). Analysis of milk quality and its importance for milk processors. Scientific Papers Animal Science and Biotechnologies. 42(1): 501-501.

Renes Carreño, E., Escribá Bárcena, A., Catalán González, M., Álvarez Lerma, F., Palomar Martínez, M., Nuvials Casals, X. and Montejo González, J.C. (2022). Study of risk factors for healthcare-associated infections in acute cardiac patients using categorical principal component analysis (CATPCA). Scientific Reports. 12(1): 28. https:/ /doi.org/10.1038/s41598-021-03970-w.

Sankhyan Varun, Thakur Y.P, Katoch Sanjeet, Dogra P.K, Thakur Rakesh (2017). Morphological structuring using principal component analysis of Rampur-Bushair sheep under transhumance production in western Himalayan region, India. Indian Journal of Animal Research. 52(6): 917-922. doi: 10.18805/ijar.B-3296.

Torres-Cárdenas, V., Torres, JOS., Melo, JM., Fuentes, NF., Pérez, AB., Calero, CAM., (2021). Application of categorical principal component analysis in the study of ovine production systems in Ciego de Ávila province, Cuban Journal of Agricultural Science. 55(4): 347-359.

TÜİK, Türkiye İstatistik Kurumu. (2022). Hayvancılık İstatistikler. https://brun.tuk.gov.tr/medas/?locale=tr. Access: Aralık, 2023.

Ulusal Süt Konseyi, (2021). Süt Raporu, Dünya ve Türkiye'de Süt Sektör İstatistikleri. Ankara, 100s. https://ulusalsutkonseyi. org.tr/wp-content/uploads/2021-Sut-Raporu.pdf.