

Paper type: Original Research

Statistical modeling of the latent variable of milk production performance in the Murciano-Granadina goats by confirmatory factor analysis

Morteza Mokhtari^{1,*}, Arsalan Barazandeh¹, Zahra Roudbari¹, Najmeh Kargar Borzi², Moslem Moghbeli Damane¹

¹Department of Animal Science, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

²Animal Science Research Department, Kerman Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education & Extension Organization (AREEO), Kerman, Iran

*Corresponding author,
E-mail address:
msmokhtari@ujiroft.ac.ir

Received: 07 Apr 2026,
Received in revised form: 30 Apr
2026,
Accepted: 05 May 2026,
Published online: 14 May 2026,
© The authors, 2027.

ORCID

Morteza Mokhtari
0000-0002-3971-628X
Arsalan Barazandeh
0000-0001-7716-7624
Zahra Roudbari
0000-0002-3357-3126
Najmeh Kargar Borzi
0000-0001-9907-1877
Moslem Moghbeli Damane
0000-0002-1853-8103

Abstract In the present study, 15,108 daily records of milk yield (DMY), milk fat percentage (DFP), milk protein percentage (DPP), and milk somatic cell score (DSCS) in the first two lactation periods of the Murciano-Granadina goats, collected from 2017 to 2024 in the southern part of Kerman province of Iran, were used. By applying latent variable modeling technique and confirmatory factor analysis a latent variable of daily milk production performance (DMP) in the first two lactation periods of the Murciano-Granadina goat breed was constructed and evaluated statistically by four goodness of fit measures including standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), Tucker-Lewis Index (TLI), and comparative fit index (CFI). The values of CFI, TLI, RMSEA, and SRMR, were 0.97, 0.92, 0.05, and 0.02, respectively, implying the suitability of the confirmatory factor model proposed for DMP latent variable in the present study. Standardized factor loadings of the traits used for describing DMP were statistically significant ($P < 0.01$) values of 0.24, -0.55, -0.54, and -0.37 for DMY, DFP, DPP, and DSCS, respectively. Therefore, it may be concluded that increase in milk yield would increase DMP. On the other hand, increases in milk fat percentage, milk protein percentage, and milk somatic cell score would decrease DMP. For constructing the latent variable of DMP in the Murciano-Granadina goats, emphasis should be on increasing milk quantity and decreasing milk composition traits including milk fat percentage, milk protein percentage, and milk somatic cell score. Pearson's correlations among the investigated measurable traits were statistically significant ($P < 0.01$), ranging from -0.14 (DMY-DFP and DMY-DSCS) to 0.30 (DFP-DPP). Pearson's correlations between the DMP latent variable and the measurable milk yield and composition traits were also statistically significant ($P < 0.01$) and were -0.76 (DFP-DMP), -0.75 (DPP-DMP), -0.52 (DSCS-DMP), and 0.33 (DMY-DMP). Overall, the study provides a statistical framework for describing DMP in contexts such as phenotypic evaluations, where a latent construct can capture the concept of milk yield and composition traits more effectively than individual traits alone.

Keywords: confirmatory factor analysis, goat, latent variable, milk production traits

Introduction

The efficiency of milk production may be influenced by several traits, such as the milk yield, milk composition

characteristics like fat and protein percentages, and milk somatic cell score (Godden et al., 2001). These characteristics are the main traits considered in breeding

programs of dairy goats (Menendez-Buxadera et al., 2010) and imply different aspects of milk production efficiency in dairy cows. The value of a selection candidate as a potential parent for the next generation is often evaluated based on multiple traits, such as growth, body conformation, and production (Gianola and Sorensen, 2004). Multivariate models perform an important role in quantitative genetics and simultaneously analyze several traits. These models also provide a deeper understanding of the biological background underlying complex, interrelated phenotypes, which is crucial for both theoretical advancements and practical applications in animal genetics (Gianola and Sorensen, 2004). Under multivariate models the number of parameters increases with the number of traits included; this parameter inflation can compromise the feasibility of statistical analyses (Silva et al., 2021) and alter the perceived genetic potential of traits, affecting selection decisions (Lynch and Walsh, 1998).

The application of statistical techniques, like structural equation modeling (SEM), for handling such concerns was introduced by Wright (1921). SEM is a multivariate statistical technique that allows for modeling relationships between latent and observed variables, enabling quantitative evaluation of theoretical or measurement models (Schumacker and Lomax, 1996). Latent variable modeling (LVM) is one of the applications of SEM (Schumacker and Lomax, 1996) and as a dimension-reduction technique, can be applied to phenotypic data to reduce dataset dimensionality and computational complexity in such a way that several phenotypes are combined to represent a few underlying latent variables (Leal-Gutierrez et al., 2018). Statistically, latent variables are constructs that cannot be directly measured but can be inferred from multiple observed phenotypes (Schumacker and Lomax, 1996). LVM provides a powerful approach for investigating biologically complex phenomena by reducing data dimensionality, allowing many phenotypes to be combined into a smaller set of underlying concepts of interest (Leal-Gutierrez et al., 2018).

Goats are known for their ability to adequately adapt to various environmental conditions and production systems that may be undesirable for other livestock species (Oliveira et al., 2016). The Murciano-Granadina is a well-known dairy goat breed in Spain that has been exported to several countries (Martinez et al., 2010).

In the present study, the application of LVM for constructing a latent variable, namely daily milk production performance (DMP), from four measurable traits of daily milk yield (DMY), daily milk fat percentage (DFP), daily milk protein percentage (DPP), and daily milk somatic cell count (DSCC) in first two lactation periods of the Murciano-Granadina goat breed, was considered. Furthermore, phenotypic correlations of the latent trait DMP with the corresponding measurable traits of DMY, DFP, DPP, and DSCC were also investigated.

Materials and methods

Data source and the studied traits

In the present study, 15,108 daily records of DMY, DFP, DPP, and DSCC in the first-two lactation periods of the 3,828 Murciana-Granadina goats, collected from 2017 to 2024 in the southern part of Kerman province of Iran, were used. The DSCC values were right-skewed. Therefore, to obtain values that are normally distributed, DSCC values were transformed into daily somatic cell score (DSCS) values through the following transformation (Valencia-Posadas et al., 2022):

$$DSCS = \log_2 \left(\frac{DSCC}{100000} \right) + 3$$

Descriptive statistics of the dataset used is shown in Table 1.

Table 1 Descriptive statistics for the measurable traits of the daily milk yield and composition traits over the first two lactation periods of the Murciano-Granadina goat breed

Trait ¹	Mean	S.D.	C.V. (%)
DMY (kg)	1.95	0.83	42.56
DFP (%)	4.94	1.33	26.92
DPP (%)	3.42	0.58	16.96
DSCS	5.07	1.82	35.90

¹DMY: daily milk yield, DFP: daily milk fat percentage, DPP: daily milk protein percentage, DSCS: daily somatic cell score

Latent variable modeling

The hypothesized theoretical model for constructing the DMP latent variable from DMY, DFP, DPP, and DSCS traits is depicted in Figure 1. The statistical technique of confirmatory factor analysis (CFA) is employed to identify the underlying latent factors that influence the variations in DMP. The CFA is one of the methods within SEM that can be applied to reduce the dimensionality of phenotypic data (Yu et al., 2019) and is employed to test the hypothesis that a set of observed variables is associated with a smaller set of latent variables (Momen et al., 2021). The following CFA model was used:

$$\mathbf{y} = \mathbf{A}\mathbf{b} + \boldsymbol{\varepsilon}$$

In which, $\mathbf{y}$ is the vector of measured traits including DMY, DFP, DPP, and DSCS in individuals, and $\mathbf{b}$ is the vector of latent factors. $\mathbf{A}$ matrix includes factor loadings that relate these factors to the measurable traits, and $\boldsymbol{\varepsilon}$ is the residual vector. The lavaan package (Rosseel, 2012) was used for fitting the above CFA model in the R environment (R Development Core Team, 2025). Four goodness of fit measures, including standardized root mean square residual (SRMR) (Bentler, 1990), root mean square error of approximation (RMSEA) (Steiger, 1990), Tucker-Lewis Index (TLI), and comparative fit index (CFI) (Bentler, 1990), were used to assess the overall adequacy of the proposed model for constructing DMP latent variable. The final confirmed model was evaluated for bias by permutation employing the bootstrapLavaan function from the lavaan package (Rosseel, 2012) with 5000 bootstrap draws. The obtained values for the latent variable of DMP by fitting CFA followed the standard normal distribution with a

mean of zero and a standard deviation of 1. Pearson's correlations between the DMY, DFP, DPP, DSCS, and DMP traits were computed using SAS 9.4 program (SAS, 2010). Preliminary analysis revealed that Pearson's correlations between the DMY, DFP, DPP and DSCS were low to medium, ranged from -0.14 (DMY-DFP and DMY-DSCS) to 0.30 (DFP-DPP). Therefore, the assumption of independence for data was acceptably held. It should be noted that, in the present study the effect of lactation period on the studied traits was not statistically significant ($P > 0.05$). Therefore, test-day records in both lactation periods were considered together for constructing a latent variable of DMP.

Results and discussion

The constructed DMP latent variable

Figure 1 provides schematic representations of the hypothesized measurement model considered for constructing DMP. The goodness of fit measures, including CFI, TLI, RMSEA, and SRMR, were 0.97, 0.92, 0.05, and 0.02, respectively, implying the suitability of the confirmatory factor model proposed for DMP latent variable in the present study. Li et al. (2026) applied a similar methodology to define the latent variable of test-day lactation performance by applying measurable traits of test-day milk yield, test-day milk fat percentage, test-day milk protein percentage, test-day somatic cell count, and test-day milk urea nitrogen in Chinese Holstein dairy cows and pointed out that these traits described the proposed latent variable appropriately. By applying multivariate factor analysis, Cecchinato et al. (2019) constructed 12 latent variables related to the milk fatty acid profile, milk composition (milk yield, contents of milk fat, protein, casein, and lactose, and SCS), and udder health in Italian Brown Swiss cows. They also confirmed the suitability of LVM for developing the constructs of latent variables from the corresponding measurable traits.

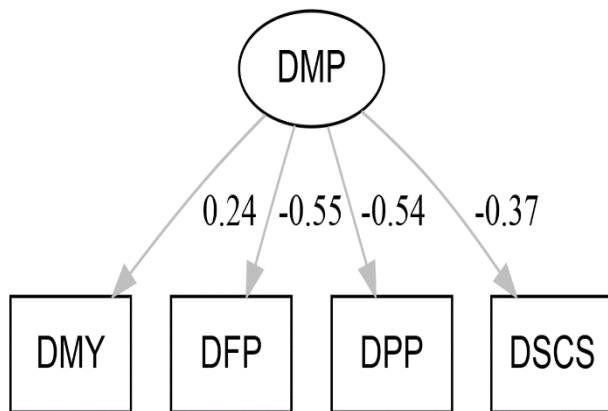


Figure 1. Representation of the latent variable for daily milk performance (DMP) and its relationships with daily milk yield (DMY), daily milk fat percent (DFP), daily milk protein percent (DPP), and daily milk somatic cell score (DSCS). For each measured variable, the estimates of factor loadings are presented.

Standardized factor loadings related to DMY, DFP, DPP, and DSCS are shown in Table 2. The standardized factor loadings can be considered as regression coefficients that show the degree of association between measured traits and the underlying latent variable (Penagaricano et al., 2015). All the standardized factor loadings were statistically significant ($P < 0.01$), implying that the measurable traits of DMY, DFP, DPP, and DSCS were appropriate indicators of the corresponding latent variable of DMP. The factor loadings were positive for DMY and negative for DFP, DPP, and DSCS (Table 2). A positive (negative) factor loading implies that an increase in any measurable variable is accompanied by an increase (decrease) in the latent variable. In the present study, the factor loading of DMY (0.24) was positive, whereas the factor loadings of other investigated traits (DFP, DPP, and DSCS) were negative. In other words, a one-unit increase in DMY is accompanied by a 0.24-unit increase in DMP and one unit increase in DFP, DPP, and DSCS is associated with -0.55, -0.54, and -0.37 unit decreases in the DMP (Table 2). Therefore, it may be concluded that for constructing the latent variable of DMP in Murciano-Granadina goats, emphasis should be on increasing milk quantity and decreasing milk composition traits including milk fat percentage, milk protein percentage, and milk somatic cell score. Li et al. (2026) constructed a latent variable of test-day lactation performance in the first parity of Chinese Holstein dairy cows and reported the standardized factor loadings for test-day milk yield, test-day milk fat percent, test-day milk protein percent, and test-day milk somatic cell score as 0.46, -0.52, -0.70, and -0.14, respectively.

Table 2 Standardized factor loadings of the traits used for describing DMP latent variable in the Murciano-Granadina goats

Trait ¹	Factor loading $\pm$ Standard error	99% confidence interval
DMY	0.24 $\pm$ 0.01	[0.21: 0.27]
DFP	-0.55 $\pm$ 0.02	[-0.60: -0.50]
DPP	-0.54 $\pm$ 0.02	[-0.59: -0.49]
DSCS	-0.37 $\pm$ 0.01	[-0.34: -0.40]

¹ DMY: daily milk yield, DFP: daily milk fat percentage, DPP: daily milk protein percentage, DSCS: daily somatic cell score

Correlations among the traits

Visual representation of the phenotypic Pearson's correlation coefficients among the investigated traits including DMY, DFP, DPP, DSCS, DMP traits are presented in Figure 2. All the correlations among the investigated measurable traits were statistically significant ($P < 0.01$), ranging from -0.14 (DMY-DFP and DMY-DSCS) to 0.30 (DFP-DPP). Low to medium correlations among these traits implying there is no concerns regarding the presence of collinearity among the investigated measurable traits including DMY, DFP, DPP, and DSCS in this study. The correlations between the DMP latent variable and the measurable milk yield and composition traits were also statistically significant ($P < 0.01$) and were -0.76 (DFP-DMP), -0.75 (DPP-DMP), -0.52 (DSCS-DMP), and 0.33 (DMY-DMP). The relatively

high correlations of DFP-DMP (-0.76) and DPP-DMP (-0.75) suggest that DFP and DPP are more strong indicators of DMP than DMY and also DSCS. Li et al. (2026) reported phenotypic correlations of test-day lactation performance, as a latent variable, in the Chinese Holstein dairy cows with test-day milk yield,

test-day milk fat percentage, test-day milk protein percentage, and test-day milk somatic cell score as 0.40, -0.59, -0.85, and -0.07, respectively, which were in general agreement with those of obtained in the present study.

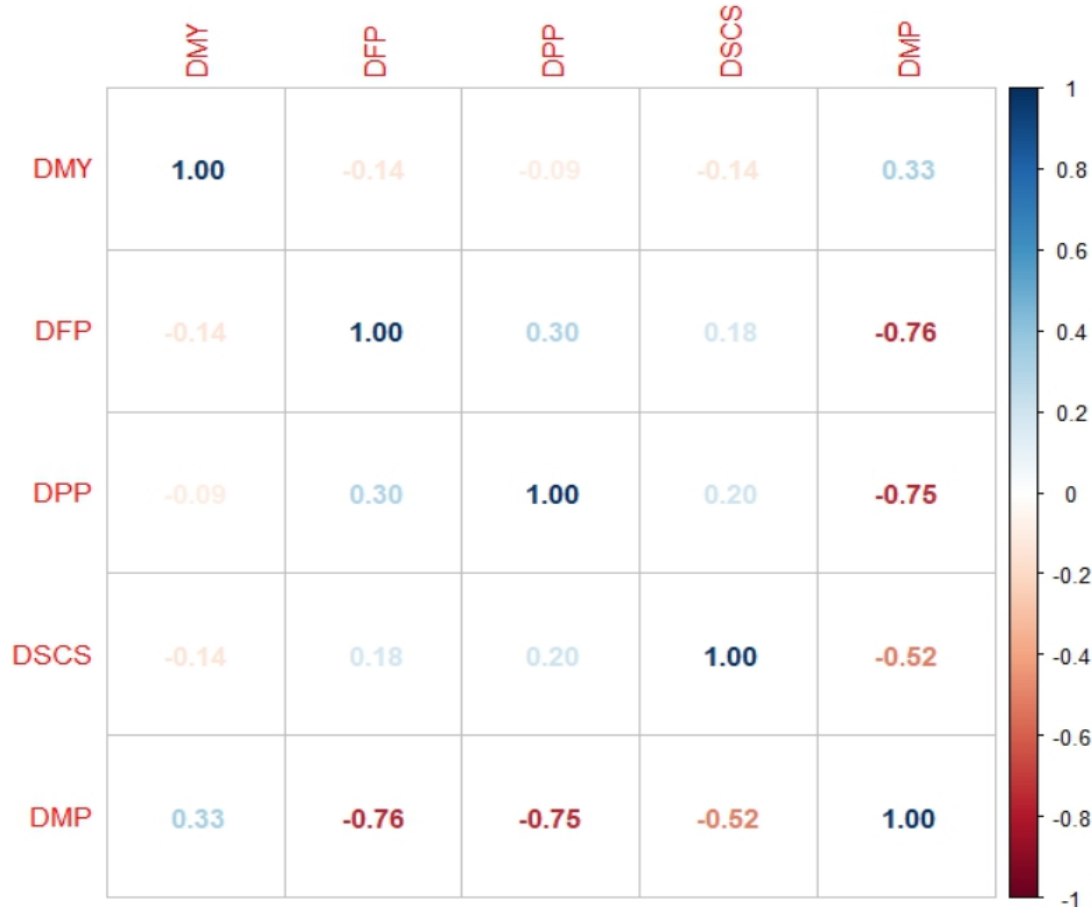


Figure 2. Visual representation of the phenotypic Pearson's correlation coefficients among the investigated traits including DMY (daily milk yield), DFP (daily milk fat percentage), DPP (daily milk protein percentage), DSCS (daily somatic cell score) and DMP (daily milk performance) traits

In the present study, the estimate of Pearson's correlations between DMP and DMY were positive, but those between DMP and other investigated traits related to milk composition (DFP, DPP, and DSCS) were negative. A positive relationship between DMP and DMY was also confirmed by an associated positive factor loading (Table 2). On the other hand, negative Pearson's correlations between DMP and DFP, DPP, and DSCS were confirmed by the corresponding negative factor loadings (Table 2). The highest negative factor loading (-0.55) and the strongest negative Pearson's correlation (-0.76) of DMP are with DFP, implying DMP is almost inversely and highly related to milk fat percentage in the first two lactations of the Murciano-Granadina goat. The lowest, in magnitude, factor loading (0.24) and the weakest Pearson's correlation (0.33) of DMP is with DMY, suggesting DMP is less influenced by the milk yield in the first two lactation periods of the Murciano-Granadina goat breed.

Conclusion

Statistical evidence revealed the adequacy of the application of LVM for constructing the latent variable of DMP in the first two lactation periods of the investigated population of Murciano-Granadina goat breed in the southern area of Kerman province of Iran. The investigated measurable milk yield and composition traits generally provided an appropriate representation of the construct developed for DMP. From a statistical perspective, these findings suggested that DMP can be regarded as an integrative measure to evaluate milk yield and composition in the Murciano-Granadina goat breed in a theoretical framework. Overall, the study provides a statistical framework for describing DMP in contexts such as phenotypic evaluations, where a latent construct can capture the concept of milk yield and

composition traits more effectively than individual traits alone. But the validation of the proposed construct for DMP needs further investigation in other herds of this breed under different production systems.

Acknowledgements

The authors would like to thank the staff of Ghale-Ganj dairy farm and Fajr Isfahan Agricultural and Livestock Company for their cooperation in data provision. This research was supported by the University of Jiroft under grant NO: 2812-04-09-335065.

Conflict of interest

The author declares no conflicts of interest between the author and other people, institutions, or organizations.

References

- Bentler, P.M., 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107, 238-246.
- Cecchinato, A., Masciotra, N.P.P., Mele, M., Tagliapietra, F., Schiavon, S., Bittante, G., Pegolo, S., 2019. Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. *Journal of Dairy Science* 102, 5254-5265.
- Gianola, D., Sorensen, D., 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* 167, 1407-1424.
- Godden, S.M., Lissemore, K.D., Kelton, D.F., Leslie, K.E., Walton, J.S., Lumsden, J.H., 2001. Relationships between milk urea concentrations and nutritional management, production, and economic values in Ontario dairy herds. *Journal of Dairy Science* 84, 1128-1139.
- Leal-Gutierrez, J.D., Rezende, F.M., Elzo, M.A., Johnson, D., Penagaricano, F., Mateescu, R.G., 2018. Structural equation modeling and whole-genome scans uncover chromosome regions and enriched pathways for carcass and meat quality in beef. *Frontiers in Genetics*, 9, 1-13.
- Li, H., Mokhtari, M., Tian, J., Sun, G., Esmailzadeh, A., Zhao, M., Wang, X., Jin, L., Chen, L., Zhang, J., Tian, R., 2026. Modelling and genetic analysis of the latent variable of lactation performance in Chinese Holstein dairy cows. *Journal of Dairy Research*, 93 35- 41.
- Lynch, M., Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits*. Oxford University Press, Sunderland, MA: Sinauer Associates, Inc.
- Martinez, A.M., Vega-Pla, J.L., Leon, J.M., Camacho, M.E., Delgado, J.V., Ribeiro, M.N., 2010. Is the Murciano-Granadina a single goat breed? A molecular genetics approach. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia* 62, 1191-1198.
- Menendez-Buxadera, A., Molina, A., Arrebola, F., Gil, M.J., Serradilla, J.M., 2010. Random regression analysis of milk yield and milk composition in the first and second lactations of Murciano Granadina goats. *Journal of Dairy Science*, 93 2718-2726.
- Momen, M., Bhatta, M., Hussain, W., Yu, H., Morota, G., 2021. Modeling multiple phenotypes in wheat using data-driven genomic exploratory factor analysis and Bayesian network learning. *Plant Direct* 5, e00304.
- Oliveira, R. R., Brasil, L. H. A., Delgado, J.V., Peguezuelos, J., León, J.M., Guedes, D. G. P., Arandas, J.K.G., Ribeiro, M.N., 2016. Genetic diversity and population structure of the Spanish Murciano-Granadina goat breed according to pedigree data. *Small Ruminant Research* 144,170-175.
- Penagaricano, F., Valente, B. D., Steibel, J. P., Bates, R. O., Ernst, C. W., Khatib, H., Rosa, G. J. M., 2015. Searching for causal networks involving latent variables in complex traits: Application to growth, carcass, and meat quality traits in pigs. *Journal of Animal Science* 93, 4617-4623.
- R Development Core Team, 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48, 1-36.
- SAS (Statistical Analysis System), 2010. SAS User's Guide, Version 9.4. SAS Institute Inc. Cary, North Carolina, USA.
- Schumacker, E.R., Lomax, G.R., 1996. *A Beginner's Guide to Structural Equation Modeling*. Erlbaum, Mahwah, NJ.
- Silva, H.T., Paiva, J.T., Botelho, M.E., Carrara, E.R., Lopes, P.S., Silva, F.F., Veroneze, R., Sterman Ferraz, J.B., Eler, J.P., Mattos, E.C., Gaya, L.G., 2021. Searching for causal relationships among latent variables concerning performance, carcass, and meat quality traits in broilers. *Journal of Animal Breeding and Genetics* 139, 181-192.
- Steiger, J. H., 1990. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 25, 173-180.
- Valencia-Posadas, M., Lechuga-Arana, A.A., Avila-Ramos, F., Shepard, L., Montaldo, H.H., 2022. Genetic parameters for somatic cell score, milk yield and type traits in Nigerian Dwarf goats. *Animal Bioscience* 35, 377-384.
- Wright, S., 1921. Correlation and causation. *Journal of Agricultural Research* 20, 557-585.
- Yu, H., Campbell, M.T., Zhang, Q., Walia, H., Morota, G., 2019. Genomic Bayesian confirmatory factor analysis and Bayesian network to characterize a wide spectrum of rice phenotypes. *Genes, Genomes Genetics* 9, 1975-1986.