ACTA SCIENTIARUM POLONORUM

Acta Sci. Pol. Zootechnica 20(3) 2021, 41–44

www.asp.zut.edu.pl pISSN 1644-0714 eISSN 2300-6145 DOI:10.21005/asp.2021.20.3.05

ORIGINAL PAPER

Received: 11.09.2021
Accepted: 14.10.2021

DETERMINATION OF THE BEST MODEL TO PREDICT MILK DRY MATTER IN HIGH MILK YIELDING DAIRY CATTLE

Burcu Kurnaz[1], Hasan Önder[1], Dariusz Piwczyński[2], Magdalena Kolenda[2], Beata Sitkowska[2]

[1]Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Turkey
[2]Department of Animal Biotechnology and Genetics, Faculty of Animal Breeding and Biology, Bydgoszcz University of Science and Technology, 85-084 Bydgoszcz, Poland

ABSTRACT

This study was aimed to determinate the best model to predict milk dry matter in high milk yielding dairy cattle. Level of milk dry matter (MDM) (%) is of great importance. The material of this study consisted of 2208 milking records of dairy cattle yielding more than 40 l per day from Polish Holstein Friesian population. In this study to estimate the milk dry matter, regression of daily milk yield (MY) (l), milk urea (MU), milk protein (MP) (%) and milk fat (MF) (%) as explanatory variables were used. To estimate the best fitting, curve estimation was used. Estimation of the curves showed that milk urea was cubic, milk yield, milk protein and milk fat were quadratic. To avoid multicollinearity where VIF value greater than 10, stepwise variable selection procedure was used. After variable selection the regression equation was obtained as $MDM = 2.879 + 1.290MF + 2.395MP - 0.039MF^2 - -0.225MP^2$ with 0.946 coefficient of determination. Our results showed that milk fat (%) and milk protein (%) can be used to estimate the milk dry matter (%) with a great achievement in high milk yielding dairy cattle.

Key words: Regression, Curve estimation, Milk component, Estimation

INTRODUCTION

The total components in milk other than water are called dry matter. Other component distinctions; milk is non-fat dry matter and fat in dry matter. The chemical composition of milk significantly affects its nutritional value. In addition, some microorganisms can be found in milk. These affect the chemical reactions and sensory qualities of milk (sut.agri.ankara.edu.tr, access date: 27.08.2021). Milk dry matter content is important in terms of giving an idea about the nutritional value of milk and processing it into dairy products [Kurt et al. 1996]. Dry matter provides efficiency from milk and dairy products. About 27% of the dry matter consists of nitrogenous substances and again about 29% is milk fat. Lactose has the largest share, with a value of approximately 37% in dry matter. Mineral substances are present at a rate of 5.95% in the total dry matter. Other substances in the composition of cow's milk are not important in terms of quantity. However, they have great importance in terms of their functions [Metin and Kaşıkçı 2010].

The purpose of the multiple regression model is to explain the relationship between a dependent (response) and independent (explanatory) variables.

In regression analysis, the contribution of some of the independent variables to the model may be insignificant. Therefore, it is necessary to determine the independent variables that will explain the dependent variable in the "best fit" way and to exclude unimportant variables from the model. This process is called "variable selection" [Alpar 2003].

Many factors affect a factor that is considered as a dependent variable in animal science. However, many researchers are hesitant about how many or which of the factors affecting a factor should be included in the regression model. For this reason, the issue of which in-

burcu2039@hotmail.com

dependent variables will be included in the model in the regression analysis is of great importance in animal science [Berberoğlu 2002].

In this study it was aimed to determinate the best model to predict milk dry matter in high milk yielding dairy cattle with using milk yield, milk urea, milk protein, and milk fat.

MATERIAL AND METHODS

The material of this study consisted of 2208 milking records of dairy cattle yielding more than 40 l per day from Polish Holstein Friesian population. Before the pruning of the data the descriptive statistics (Mean ±StdDev) of the variables MDM, MY, MF, MP, and MU were estimated 12.94 ±1.01, 28.58 ±9.02, 3.99 ±0.80, 3.42 ±0.40, and 249.03 ±71.96, respectively with number of 29942 number of records, parities were 1 and 2.

To estimate the milk dry matter (MDM) (%), daily milk yield (MY) (l), milk urea (MU) (mg · 100 g^{-1}), milk protein (MP) (%) and milk fat (MF) (%) as explanatory variables were used. Descriptive statistics (Mean ±StdDev) of the variables MDM, MY, MF, MP, and MU were estimated 12.23 ±0.54, 43.83 ±3.39, 3.52 ±0.46, 3.15 ±0.19, and 264.89 ±66.54, respectively.

Multiple linear regression method was used to estimate the model parameters in the first stage of the study [Carl and Kühn 2008]. Curve estimation was applied to the explanatory variables on dependent MDM variable to understand the fitting curve shape [Weisberg 2005]. Stepwise variable selection procedure was used to avoid multicollinearity problem that decrease the reliability of the model with VIF value greater than 10 [Tırınk et al. 2020] and simplification of the model [Önder and Cebeci 2002]. All statistical analysis were executed using SPSS V.20.0 with license of Ondokuz Mayıs University (IBM SPSS Statistics, Armonk, NY, USA).

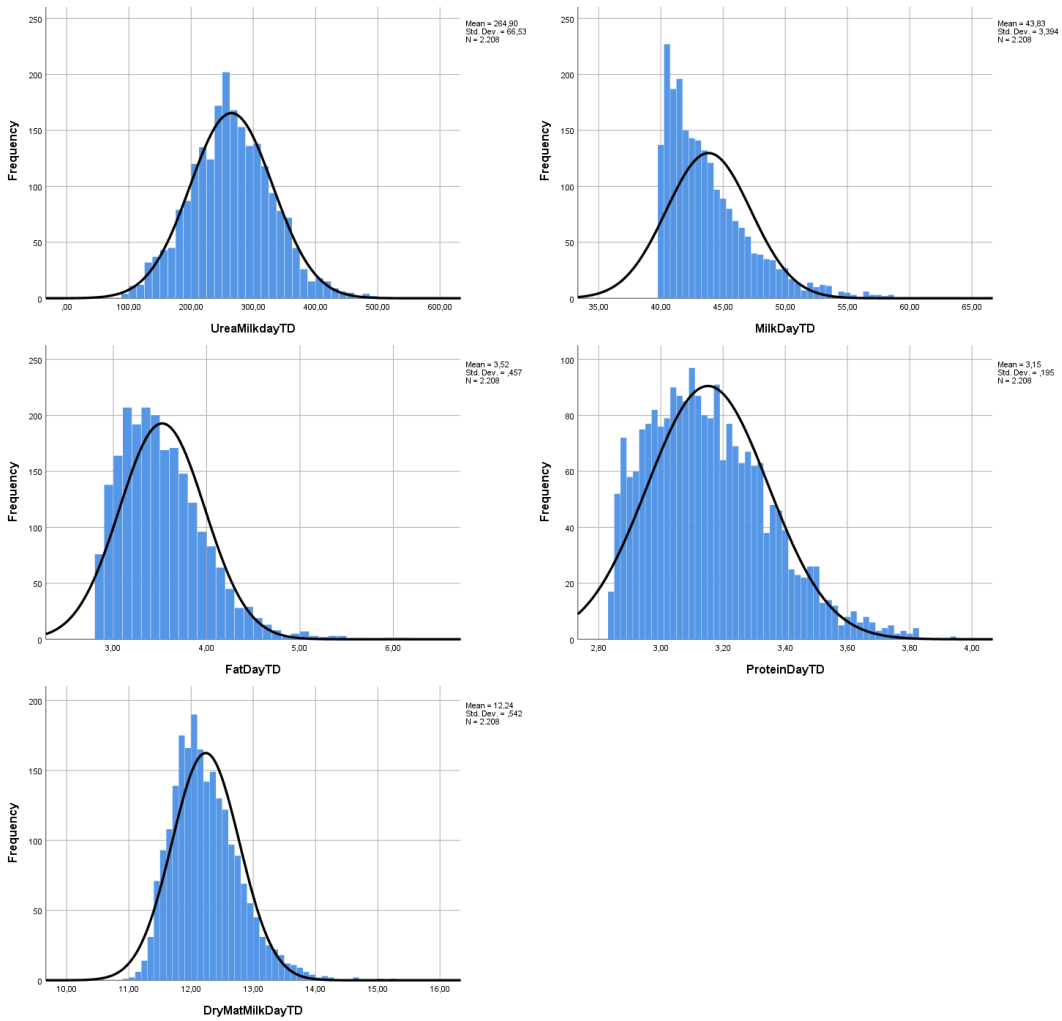

Fig. 1. Distribution of the variables

RESULTS AND DISCUSSION

When the distribution of the variables were examined by Kolmogorov-Smirnow One Sample test, MU, MF, MP, and MDM variables had Normal distribution. But MY variable was out of Normal distribution. Situation of the MY variable was came from selection of the milk yield over than 40 l per day (Fig. 1).

Obtained estimation equation of multiple linear regression was:

$$MDM = 5.792 - 6.4 \cdot MU - 0.001 \cdot MY + \\ +0.992 \cdot MF + 0.946 \cdot MP$$

This model was statistically significant (F = 9361.206; P < 0.001) and the adjusted coefficient of determination (R^2) was observed as 0.944. The significance and adjusted R^2 values showed that this equation can be used to estimate the milk dry matter. But some regression coefficients such as for MU and MY variables were not significant. In that situation curve estimation was used to fit the explanatory variables on dependent variable. Curve estimation results were given in Table 1.

Table 1. Curve estimation R^2 values of the explanatory variables on milk dry matter

	Linear	Quadratic	Cubic	Logarithmic
MU	0.00698	0.00911	0.00915	0.00817
MY	0.01584	0.01719	0.01719	0.01630
MF	0.83464	0.83504	0.83504	0.83076
MP	0.28285	0.28913	0.28934	0.27924

According to the curve estimation results, MU was determined as cubic, MY was determined as quadratic, MF and MP variables were determined as quadratic. The estimation equation of complete model was:

$$MDM = 3.418 + 0.002MU - 0.026MY + \\ +1.276MF + 2.339MP - 3.411 \cdot 10^{-6}MU^2 + \\ +1.291 \cdot 10^{-9}MU^3 + 0.276 \cdot 10^{-3}MY^2 - \\ -0.37MF^2 - 0.217MP^2$$

This model was also statistically significant (F = 4853.294; P < 0.001) and adjusted R^2 value was observed as 0.946. But the model still had some unimportant variables that decreased the model reliability. In that case forward stepwise variable selection procedure was applied to determine the best model. After variable selection obtained estimation equation was;

$$MDM = 2.879 + 1.290MF + 2.395MP - \\ -0.039MF^2 - 0.225MP^2$$

The last model was found statistically significant (F = 9576.156; P < 0.001) and adjusted R^2 value was observed as 0.946. This model had no insignificant variable and free from multicollinearity problem with VIF values smaller than 10. The residuals had normally distributed with mean of 0.00 and variance of 0.016. The last mentioned equation can be robustly used to estimate milk dry matter.

CONCLUSION

To estimate the milk dry matter in percentage, milk fat and milk protein percentages in quadratic manner can be used with a high robustness for high milk yielding dairy cattle. In this study statistical procedures were described step by step to give some opinion to the young researchers.

REFERENCES

Alpar, R. (2003). Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. 2. Baskı, Nobel Basımevi, Ankara, Turkey.

Berberoğlu, E. (2002). Application of Model Determination to Animal Science Using the Variable Selection Method. MSC thesis, Gaziosmanpaşa Univ, Tokat, Turkey.

Carl, G., Kühn, I. (2008). Analyzing spatial ecological data using linear regression and wavelet analysis. Stochastic Environmental Research and Risk Assessment, 22, 315–324. DOI: 10.1007/s00477-007-0117-2.

Kurt, A., Çakmakçı, S., Çağlar, A. (1996). Süt ve mamülleri analiz metotları rehberi, Atatürk Univ Press, pp: 18, Erzurum, Turkey.

Metin, M., Kaşıkçı, M. (2010). Milk Component and its Proccess. 9. Edition. Ege Univ Engineering Faculty Press, 33, pp:182, İzmir, Turkey.

Önder, H., Cebeci, Z. (2002). Variable Selection in Logistic Regression. Çukurova Üniversitesi Ziraat Fakültesi Dergisi, 17(2), 105–114.

Tırınk, C., Abacı, SH., Önder, H. (2020). Comparison of Ridge Regression and Least Squares Methods in the presence of multicollinearity for body measurements in saanen kids. Journal of the Institute of Science and Technology, 10(2), 1429–1437. DOI: 10.21597/jist.671662.

Weisberg, S. (2005). Applied Linear Regression. Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN 0-471-66379-4. DOI: 10.1002/0471704091.

OKREŚLENIE NAJLEPSZEGO MODELU DO PRZEWIDYWANIA SUCHEJ MASY MLEKA U WYSOKOWYDAJNEGO BYDŁA MLECZNEGO

STRESZCZENIE

Niniejsze badanie miało na celu określenie najlepszego modelu służącego do przewidywania suchej masy mleka u wysokowydajnego bydła mlecznego. Poziom suchej masy mleka (MDM, %) ma kluczone znaczenie dla wydajności. Materiał do badań stanowiło 2208 rekordów udojowych bydła mlecznego o wydajności powyżej 40 l/dobę z populacji krów rasy polska holsztyńsko-fryzyjska. Do oszacowania suchej masy mleka wykorzystano jako zmienne objaśniające regresję dziennej wydajności mleka (MY, l), zawartość mocznika w mleku (MU), białka (MP, %) i tłuszczu (MF, %). Do oszacowania najlepszego dopasowania wykorzystano estymację krzywej. Oszacowanie wykazało, że krzywa mocznika była sześcienna, natomiast krzywe wydajności, białka i tłuszczu były kwadratowe. Aby uniknąć wielokoliniowości, gdy wartość VIF jest większa niż 10, zastosowano procedurę stopniowego wyboru zmiennych. Po doborze zmiennej otrzymano równanie regresji jako $MDM = 2,879 + 1,290MF + 2,395MP - -0,039MF^2 - 0,225MP^2$ ze współczynnikiem determinacji 0,946. Nasze wyniki wykazały, że zawartość tłuszczu (%) i białka (%) w mleku można z dużym powodzeniem wykorzystać do oszacowania suchej masy mleka (%) u bydła mlecznego o wysokiej wydajności.

Słowa kluczowe: regresja, dopasowanie krzywej, składniki mleka, estymacja

Burcu Kurnaz https://orcid.org/0000-0001-5613-6992
Hasan Önder https://orcid.org/0000-0002-8404-8700
Dariusz Piwczyński https://orcid.org/0000-0001-8298-2316
Magdalena Kolenda https://orcid.org/0000-0002-9260-4391
Beata Sitkowska https://orcid.org/0000-0002-1036-7450