Determination of Factors Affecting Mastitis in Holstein Friesian and Brown Swiss by Using Logistic Regression Analysis**

Yasin Altay[1,*], Büşra KILIÇ[2], İbrahim AYTEKİN[3], İsmail KESKİN[3]

[1] Osmangazi University, Faculty of Agriculture, Deparment of Animal Science, Eskişehir, Turkey
[2] TİGEM, Aksaray, Turkey
[3] Selcuk University, Faculty of Agriculture, Deparment of Animal Science, Konya, Turkey

ARTICLE INFO

ABSTRACT

The aim of this study was to determine subclinical mastitis with the help of logistic regression of milk quality determined factors and some features the research material consisted of 204 (145 Holstein, 59 Brown Swiss) dairy cattle raised in a private cattle farm in Konya Province, Turkey. The independent variables considered for the detection of subclinical mastitis are breed, somatic cell number (SCC), color values (L, a, b, H, C), freezing point (FP), pH, electrical conductivity (EC), milking day (MD), lactation order (LO). The dependent variable of logistic regression was CMT score. According to the results of the study, the spesicity was 95.7% and the sensitivity was 57.6%. In general, the predicted value of the accuracy of all data was 83.3%.

1. Introduction

Mastitis is an udder disease that causes significant economic losses in dairy cattle holdings (Duval, 1969; Aytekin, 2014; Şahin, 2014). It is estimated that 20-30% of the economic losses of mastitis are due to clinical mastitis, while the remaining losses are estimated to be caused by subclinical mastitis (Tekeli, 2005; Mammadova, 2013). The main objective of dairy cattle enterprises is to make a profit and increase productivity (Boztepe, 2015). For this reason, avoiding the average milk loss due to mastitis and predicting the risks that may occur both increases the profitability in the enterprise and provides great benefit in taking precautions against an undesired possible situation.

Regression analysis investigates the relationship be- tween independent variable or variables and dependent variable; can be simple or multiple and can be applied after the assumption of linearity, normality, homogene ity, summability (Menard, 2002). The de-

pendent variable is the variable that is explained or estimated in the regression model, and this variable is assumed to be related to the independent variable or variables. The dependent variable must be continuous (data obtained from measurement, weighing or analysis). However, in some cases, it may be encountered that the dependent variable is discrete (data obtained by counting). In such cases, the relationship between dependent and independent variables can be examined with the help of logistic regression. The independent variable is the explanatory variable in the regression model and is used to estimate the value of the dependent variable (Korkmaz, 2012). In recent years, logistic regression analysis has been widely used in social, health and natural sciences with the use of statistical package programs (Bircan, 2004; Akın, 2018).

The aim of this study is to determine the relationship between the presence of mastitis disease and milk quality factors of the Holstein and Brown Swiss cattle during different lactation in a private enterprise in Konya using logistic regression analysis.

---

* Corresponding author email: *yaltay@ogu.edu.tr*

**Short communication

2. Materials and Methods

In this study, the research material consisted of 204 (145 Holstein, 59 Brown Swiss) dairy cattle raised in a private cattle farm in Konya Province, Turkey. There are 95 cows in the 1st lactation, 38 in the 2nd lactation, 27 in the 3rd lactation, 9 in the 4th lactation, 22 in the 5th lactation and 13 in the 6th and above lactation.

The California mastitis test was first described in 1957 and quickly and easily detects the disease. (Schalm and Noorlander, 1957; Sanford et al., 2006). The presence of mastitis in the dairy cattle was determined by the California Mastitis Test (CMT) and the factors thought to be effective in the presence of mastitis in cows (breeding, somatic cell number (SCC), color values (L, a, b, H, C), freezing point (FP), pH, electrical conductivity (EC), milking day (MD), lactation order (LO)) were examined by logistic regression analysis. According to the CMT score, cows with mastitis were coded as 1 and cows without mastitis were coded as 0 and analyzes were performed accordingly.

As a result of logistic regression analysis, risk factors affecting subclinical mastitis were determined. In logistic regression analysis, the model is as follows.

$$P(Y=1/X=x) = [1 + e^{-(b0 + b1 X1 + b2 X2 + ... + bp Xp)}]^{-1}$$

where,

$P(Y=1/X=x)$ = Probability of Y = 1 when X = x

(X independent variable probability of mastitis when it receives x value)

b_0 = Constant of regression

e = Natural logarithm

Linear and nonlinear relationships between dependent and independent variables can be estimated by regression analysis. In simple terms regression, the dependent variable is obtained from different data types (continuous and discrete) and linear and nonlinear regressions according to the status of the relationship with independent variables (linear, quadratic, cubic, etc.), and one and more independent variables means simple and multiple regression.

The basis of logistic regression also called logit models, is based on odds raito. Odds ratio compares the likelihood that an event will occur and the probability that it will not. In linear models, the logistic regression is obtained by taking natural logarithm because probability ratios cannot be included in the model. The maximum likelihood, which is the most commonly used method, was used in the parameter estimation of the logistic regression model. The evaluation of the data was obtained using SPSS version 18 (SPSS Inc., IL, USA).

Logistic regression, which is one of the nonlinear regression methods, is widely used since it is not affected by the assumptions of linear models. If the dependent variable shows poisson distribution as binary (0-1), it means logistic regression. In other words, the expected value of the dependent variable according to independent variables is to calculate as the probability (Tatlıdil, 1996). Logistic regression tries to estimate the probability of taking the value of 1 instead of estimating the dependent variable (Alpar, 2011). Since the results are a probability value, they take values between 0-1 (Eyduran, 2005). As predicted by linear regression coefficients, in estimating regression coefficients, weighted least squares are obtained by maximum likelihood method and discriminate function rather than least squares method (Eyduran, 2008). The significance test of the regression coefficients in the model can be examined with the help of the Likelihood Ratio, Wald and Score (Lagrange multiplier) tests (Alpar, 2011). In addition, regression coefficients are important if the odds ratio includes a confidence interval value of 1.

3. Results and Discussion

In the logit model for the detection of mastitis, a base model was created in the first stage and an enter method was used in which all the independent variables were combined. When the base model was examined, 138 animals without mastitis were correctly estimated and the correct classification rate was 100%. All 66 specimens with mastitis were incorrectly estimated, the correct classification rate was 0% and the correct classification rate of 204 cows was 67.6%. Maximum likelihood method was used in the analysis of the logit model, in which all independent variables were considered. The significance control of the model was checked by the Chi-square test and was found to be statistically significant (P <0.05).

In analyzing the fit of the model, L (likelihood) is likely to be estimated by independent variables and L + (- 2 Log likelihood) = 1. In this case, the -2 Log likelihood value decreases the fit of the model increases. If -2 Log likelihood is 0, the model is perfectly compatible. -2 Log likelihood value of the study model was found to be 161.749.

Another criterion of the fit of the model is Cox - Snell R^2 and Nagelkerke R^2 value. According to the results of the analysis, Cox - Snell R^2 and Nagelkerke R^2 values were used to determine the percentage change of the independent variables in the dependent variable, and the values were 38.6% and 53.9% respectively. The reason for the low coefficients of determination in the logit model is thought to be due to the fact that the range of variation of the independent variables considered is too high. It is also an indication that the presence of other factors affecting the dependent variable is undeniable.

The established model for the detection of mastitis is P(y) = [1+ e-(-6.744 + 0.0000109*SCC + -0.140*L+ -0.672*a + 1.527*b + 0.025*H + -1.452*C + -10.724*FP + 0.092*pH + 2.026*EC + 0.001*MD +0.799*Breed(1) + 0.532*L0(1) + 0.690*L0(2) + 1.190*L0(3) - 0.605*L0(4) 0.441*L0(5))]-1].

After estimates made with the help of this model, of the 138 animals without mastitis (0), 132 (0) and 6 (1) were found to have mastitis. Of the 66 animals with mastitis (1), 38 were mastitis (1) and 28 were healthy (0).According to the results of the study, the spescifity was 95.7% and the sensitivity was 57.6%. In general, the predicted value of the accuracy of all data was 83.3%..

Table 1
Logistic regression parameter estimates and odds ratios results

Variables	B	S.E	Wald	df	Sig.	Exp(B)	95 % CI for Exp (B)	
							Lower	Upper
Step 1[a] Breed(1)	0.799	0.562	2.023	1	0.155	0.450	0.150	1.352
SCC	0.0000109	0.0000023	19.934	1	0.000	1.000	1.000	1.000
L	-0.140	0.206	0.463	1	0.496	0.869	0.581	1.301
a	-0.672	1.174	0.327	1	0.567	0.511	0.051	5.100
b	1.527	2.043	0.559	1	0.455	4.606	0.084	252.560
H	0.025	0.025	0.237	1	0.626	1.026	0.926	1.135
C	-1.452	2.161	0.451	1	0.502	0.234	0.003	16.181
FP	-10.724	9.217	1.354	1	0.245	0.000	0.000	1542.631
pH	0.092	2.798	0.001	1	0.974	1.096	0.005	263.820
EC	2.026	0.865	5.482	1	0.019	7.586	1.391	41.374
MD	0.001	0.002	0.108	1	0.742	0.999	0.996	1.003
LO			4.398	5	0.494			
LO(1)	0.532	0.579	0.845	1	0.358	1.702	0.548	5.290
LO(2)	0.690	0.664	1.079	1	0.299	1.993	0.543	7.322
LO(3)	1.190	1.018	1.366	1	0.242	3.287	0.447	24.183
LO(4)	-0.605	0.800	0.572	1	0.449	0.546	0.114	2.620
LO(5)	-0.441	0.933	0.224	1	0.636	0.643	0.103	4.006
Constant	-6.744	27.767	0.059	1	0.808	0.001		

[a] Variable enterned on step 1; Breed, SCC, L, a, b, H, C, FP, pH, EC, MD, LO.

Table 1 shows the beta coefficients (B), standard error (SE), Chi-square values according to Wald statistics (Wald), degrees of freedom (df), significance (P value), odds ratios (Exp (B)) and confidence intervals of odds rates. It was seen that SCC variable had statistically significant effect on mastitis (P <0.05). Other independent variables had no statistically significant effect (P> 0.05).

4. Conclusions

In cases where the dependent variable is discrete, it is not correct to examine the relationship between the dependent and independent variables by classical regression analysis. In such cases, one of the ways to be used is logistic regression analysis. Logistic regression analysis, classification and assignment process can be done and they do not require the assumption of normal distribution and continuity provides some advantages.

As a result of this study, cows were found to be as high as 83.3% in determining whether mastitis. It was found that SCC and EC were effective on mastitis and other independent variables had no statistically significant effects. Only the SCC and EC independent variables explain whether or not mastitis is an advantage. It should also be remembered that the accuracy of the obtained model depends on factors such as the consistency of the analyzed data and the number of variables.

5. References

Akin M, Hand C, Eyduran E, Reed BM (2018). Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. Pl. Cell Tissu. OrganCult.,132:545-559.

https://doi.org/10.1007/s11240-017-1353-x

Alpar R (2011). Applied Multivariate Statistical Methods, 3rd Edition, Detay Publishing, Ankara.

Aytekin İ, Boztepe S (2014). Somatic Cell Count, Importance and Effect Factors in Dairy Cattle. Turkish Journal of Agriculture-Food Science and Technology, 2(3), 112-121.

Bircan H (2004). Logistic regression analysis: An application on medical data. Kocaeli University Journal of Social Sciences, (8), 185-208.

Boztepe S, Aytekin İ, Zulkadir U (2015). Dairy Cattle, 1st Edition, Selcuk University Publishing, Konya.

Duval J (1969). Treating mastitis without antibiotics," EAP Publication 69, 1969.

Eyduran E, Ozdemir T, Çak B, Alarslan E (2005). Using of logistic regression in Animal Science. Applied Sci, 5(10), 1753-1756.

Eyduran E (2008). Usage of penalized maximum likelihood estimation method in medical research: An alternative to maximum likelihood estimation method. J. Res. Med. Sci, 13(6), 325-330.

Korkmaz M, Güney S, Yiğiter Ş (2012). The importance of logistic regression implementations in the Turkish livestock sector and logistic regression implementations/fields. Harran Journal of Agriculture and Food Sciences, 16(2), 25-36.

Menard S (2002). Applied logistic regression analysis (Vol. 106). Sage.

Mammadova N, Keskin İ (2013). Application of the support vector machine to predict subclinical mastitis in dairy cattle. The Scientific World Journal.

Sanford CJ, Keefe GP, Sanchez J, Dingwell RT, Barkema HW, Leslie KE, Dohoo IR (2006). Test characteristics from latent-class models of the California Mastitis Test. Preventive Veterinary Medicine 77, 96–108.

Schalm O, Noorlander D (1957). Experiments and observations leading to the development of California mastitis test. Journal of American Veterinary Medical Association 130, 199–204.

Şahin A, Yıldırım A (2014). The Mastitis Case in Water Buffalo. Turkish Journal Of Agriculture- Food Science And Technology, 3(1), 1-8.

Tatlıdil H (1996). Uygulamalı Çok Değişkenli İstatiksel Analiz. Ankara, Cem Web Ofset.

Tekeli T (2005). Mastitis: Quality milk production and somatic cell count in the process of the European Union. Guzelis Pub. Co., Konya.