ANALYSIS OF THE EFFECTS OF SOME ECOLOGICAL FACTORS ON UDDER HEALTH BY USING NONLINEAR REGRESSION MODELS

ÖZTÜRK, I.

Department of Animal Science, Biometry Genetics Unit, Agricultural Faculty, Harran University, Şanlıurfa, Turkey
(e-mail: ozirfan23@yahoo.com; phone: +90-505 665 7583; fax: phone: +90-414-318-1431)

Abstract. Mastitis is an important udder disease causing loss of yield in dairy cattle farming and environmental factors that play a significant role in the emergence of the disease. Factors, such as infectious diseases, live weight, cattle breed, nutritional status, environment and milking hygiene, which are thought to affect the presence of mastitis disease were discussed in the study. The data of the study was obtained randomly from the records of veterinarians and small and medium-sized enterprises operating in the two provinces of Turkey. Based on these results, we found that the environment and milking hygiene, and adequate and balanced nutrition would lead to a 6.2%, 11.3%, and 94.9% lower risk on mastitis, respectively. Of the other hand, it was determined that cattle with various symptoms of infectious diseases had a 4.3 times greater risk of mastitis than healthy cattle. It was concluded that taking into account the impact rates of factors affecting the risk of mastitis of enterprises, it can contribute to the prevention of milk yield losses.
Keywords: *mastitis, milking hygiene, environmental factors, Probit and Logit regression*

Introduction

Mastitis disease is one of the biggest problems that cause yield losses in milk production enterprises. It is reported that more than 130 microorganisms cause the occurrence of mastitis (Lakew et al., 2009; Tel et al., 2009). Mastitis is observed in two ways depending on the clinical course. The first is subclinical mastitis, which is not taken seriously by farmers because it is invisible to the eye and can be described as a latent disease, while the other is clinical mastitis, which is manifested by symptoms such as hardening, swelling, flushing in the udder tissue (Ndahetuye et al., 2020; Cobirka et al., 2020). Mastitis disease causes milk yield losses both directly and indirectly. The direct effect of mastitis is the milk yield loss caused by the damages in udder tissue over time. On other hand, losses that it causes in reproductive performance and increases that it causes in treatment expenses can be shown as examples of the indirect effect of mastitis (Özyurt, 2013).

Research results noted a negative correlation between mastitis and reproductive performance (Chebel et al., 2004). This may indirectly affect milk yield losses in cases where there is failure to conceive or due to delayed insemination. Guimaraes et al. (2017) reported that cattle might suffer from milk loss up to 77% due to mastitis during the lactation period. In the studies conducted in different periods in Turkey, mastitis findings showing differences between 5% and 60% by provinces were encountered (Sabuncuoğlu et al., 2003; Kılıç and Keskin, 2019). Mastitis is not only a problem in the undeveloped countries, but also always an important risk factor in modern dairy farming enterprises in developed countries. In a study conducted in the Netherlands, as a developed country on dairy cattle breeding, it was found that 38% of dairy cattle (Rougoor et al., 1999) had mastitis, while in another study conducted in dairy farming enterprises in France, it was

reported that 29% of cows had mastitis (Longo et al., 2001; Sabuncuoğlu and Çoban, 2006).

Given the large economic losses caused by mastitis disease in dairy farming enterprises, it is clear that the impact rates of the factors causing mastitis should be determined by appropriate regression models. The presence of mastitis disease discussed here and some of the factors affecting it are qualitative (categorical) variables, and some are continuous variables. Since qualitative variables indicate the presence or absence of a characteristic, the way to digitize these characteristics is to create dummy variables that take values of 0 and 1 (Gujarati and Porter, 2009; Güneri and Durmuş, 2020). In the model created in this study, like independent variables, dependent variables include Bernoulli random variables consisting of categorical data with two options (1/0).

The risk of mastitis in cattle is not only microbial, but also under the influence of some environmental factors. For example, this disease can also develop due to many environmental factors such as weather conditions, age, weight, nutrition and care status of the animal (Mekonnen et al., 2017; Jamali et al., 2018; Özenç, 2019). In order to model mastitis findings under the influence of these factors, instead of the least-squares (LS) method, it is necessary to take advantage of approaches that better explain binary dependent variables (Long and Freese, 2014; Mert, 2016). For this purpose, by using Logit and Probit regression models, it was attempted to explain the impact rates of factors that could cause the presence of mastitis disease leading to a significant loss of yield in dairy cattle. In this way, it is aimed to draw attention to the effect rates of some environmental factors that will reduce mastitis to the lowest incidence.

Materials and Methods

The research was carried out in 2020 in Adıyaman (37.763650; 38.277259, GPS: 37° 45' 49.14" N 38° 16' 38.132" E) and Şanlıurfa (37.167404; 38.795515; GPS: 37° 10 '2.654" N 38° 47' 43.854 " E) provinces taking place in the Southeastern Anatolia region of Turkey. The data were obtained randomly from some small (1 to 10 head) and medium-sized (11 to 100 head) cattle enterprises located in the rural area in question. For this purpose, the records of veterinarians serving in the region were also used. In the records, information such as cattle's earring numbers, breeds, farms in which they are located, disease symptoms, body weight, age, and artificial insemination was reached. Cattle with missing registration information and some variables were eliminated, and data obtained from the remaining 180 dairy cattle were used. By visiting the 69 enterprises where the cattle in question were registered, feeding conditions, and environment and milking hygiene of cattle were observed. In addition, as a result of a survey conducted with owners of the enterprises, factors related to the environment and milking hygiene were scored with values between 0 and 100 points. Descriptions related to the dependent and independent variables to be included in the regression model are given in *Table 1*.

The presence of mastitis in dairy cattle in the enterprises was defined as a dependent variable (Y). Whether there was mastitis (1) or not (0) was coded as a binary categorical variable. The live weight of cattle, which were considered as independent variable, was defined by X_1 and its unit was kg. Environmental hygiene (X_2) and milking hygiene (X_3) in the enterprises were scored between zero and one hundred (0-100).

Table 1. *Descriptions of dependent and independent (explanatory) variables*

Variables	Meaning of the variable	Description	Codding
Dependent (Y) Independent	Mastitis	Whether or not there is mastitis in cattle in randomly determined enterprises	Mastitis yes=0 (Reference) Mastitis no=1
(X_1)	Live weight	Live weight of the cattle (Continuous variable)	kg
(X_2)	Environmental hygiene	Hygiene status of the enterprise: scored between 0 and 100	Low score = poor hygiene High score = good hygiene
(X_3)	Milking hygiene	Whether attention is paid to milking hygiene: scored between 0 and 100.	Low score = poor hygiene High score= high hygiene
(X_4)	Infectious disease	Whether cattle have any infectious diseases (categorical variable)	No disease=0 (reference) Have disease = 1
(X_5)	Nutrition	Balanced nutritional status of dairy cattle	Malnutrition =0 (Reference) Adequate Nutrition = 1
(X_6)	Cattle breed	Breeds of cattle in the enterprises (categorical variable)	Local breed = 1(Reference) Holstein friesian = 2 Simmental = 3

A low score indicates that hygiene is poor and bad, while a high score indicates that hygiene conditions are good. Environmental and milking hygiene scores were categorized as in *Table 2*. As can be seen in *Table 2*, 20 points were given for each operation performed by farmers in terms of environmental and milking hygiene; thus, the enterprise performing all expected operations earned a score of 100.

Table 2. *Classification of scores given for environmental and milking hygiene procedure.*

Scores	Hygiene procedure	
	Environmental hygiene	Milking hygiene
1-(20%)	Dry base	Simple udder and teat cleaning
2-(40%)	Material hygiene	Application of disinfectant to teats before milking
3-(60%)	Periodic environmental disinfection	Cleaning with disposable cloths
4-(80%)	Disinfecting shoes at the entrance to the enterprise	Cleaning milk residues at the end of milking
5-(100%)	Outsiders are not allowed to enter into the enterprise and milking households	Applying teat dipping at the end of milking

Whether the cattle in question had any infectious diseases was defined by X_4. Cattle that had disease were coded by "1", while cattle that did not have disease were accepted as references and coded by "0". The nutritional status defined as X_5 was coded by "1" if the cattle were fed with an adequate and balanced ration, and by "0" if they were fed insufficiently and with haphazard feed. Cattle breeds in the enterprises (X_6) were coded as local breed (1), Holstein (2), and Simmental (3). Local breed was accepted as the reference. In the region in general, as a local breed, there is South Anatolian Red (SAR), which is adapted to hot climatic conditions.

In short, as seen in *Table 1*, the dependent variable is a binary variable, while the independent variables are categorical, binary categorical, and continuous variables. In regression model estimations to be made by LS method, problems such as the occurrence

of the problem of changing variance, non-normal distribution of residues, or the loss of the importance of the coefficient of determination (R^2) can be encountered due to the fact that the dependent variable is a binary categorical variable. Therefore, it is more convenient to use the logistic model whose parameters are estimated iteratively by the maximum likelihood method (Mert, 2016).

In Logit and Probit regression approaches, assumptions that are in the LS method are not sought. The only assumption is that the dependent variable is a binary categorical variable such as successful-unsuccessful and sick-healthy (1 and 0). For independent variables, there is not any restriction. In other words, independent variables can be nominal, ordinal, or continuous. The most basic difference between the Logit model and the Probit model is that while the Logit model uses the logistic odds ratio (the odds ratio is the probability obtained by dividing the probability of an event occurring by the probability of it not occurring), the Probit model uses the Gaussian distribution function. In logistic regression, the probability of the interested event is P_i, and the dependent variable is shown as ln ($P_i/(1-P_i)$) and it is defined as a Logit function (İnal et al., 2006; Mert, 2016).

Logit regression model

In logistic regression analysis, as in other model configuration methods, the goal is to create the best prediction model with minimum independent variables. As in linear regression, in logistic regression analysis, estimates are tried to be made based on some variables. However, in linear regression, the value of the dependent variable is tried to be estimated by using independent variables, while in logistic regression, the probability of occurrence of one of the values that the dependent variable can take is calculated (Montgomery et al., 2012).

Logistic regression is a powerful alternative method that could be used in conditions where the assumptions of other methods cannot be met. Moreover, Nusinovici et al. (2020) reported that the estimation performance of the logistic regression was at least as good as machine learning techniques in some studies. Logistic regression refers to models that reveal the shape of the relationship between weak-scale (categorical) variables.

The basis of the logistic regression model is based on the odds ratio. The odds ratio is obtained by dividing the probability of an event occurring by the probability of the event not occurring. By taking the natural logarithm of the ratio in question (odds), the parameters of the logistic regression model are estimated using the Maximum Likelihood method (Gujarati and Porter, 2009).

Signs of the coefficients belonging to the independent variables in the estimated model show the direction of the relationship. A positive coefficient indicates that the probability of occurring of an event will increase, and a negative coefficient indicates that it will decrease. Another meaning of negative coefficients is that the odds ratio is less than 1 and its probability is less than 0.5. The fact that the coefficient is zero means that the probability of occurring of an event will not change (Mert, 2016).

In binary logistic regression analysis, the functions given in *Eq.1* and *Eq.2* are used.

$$P(Y) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \qquad \text{(Eq.1)}$$

$$P(Y) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \qquad \text{(Eq.2)}$$

In case of multivariate model, Z value in *Eq.1* is written as in *Eq.3*

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \qquad (Eq.3)$$

Beta (β_i) parameters in *Eq.3*; $\beta_0, \beta_1, \beta_2, \ldots, \beta_P$ are the coefficients of the logistic regression.

Calculation of logistic regression parameter coefficients is performed as follows.

If *Q(Y)=1-P(Y)*, the odds ratio is found as:

$$\frac{P(Y)}{Q(Y)} = \frac{P(Y)}{1-P(Y)} = \frac{\frac{e^Z}{(1+e^Z)}}{1-\frac{e^Z}{(1+e^Z)}} = e^Z = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p} \qquad (Eq.4)$$

If the natural logarithm of both sides of the odds ratio equation (*Eq.4*) is taken, *Eq.5* is obtained;

$$\ln\frac{P(Y)}{Q(Y)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \qquad (Eq.5)$$

If the ratio of two separate odds values to each other is calculated, the odds ratio (OR) is obtained:

$$OR = \frac{P(Y)}{Q(Y)} = e^Z = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p} = Exp(\beta) \qquad (Eq.6)$$

In *Eq.6*, the *Exp(β)* values of each parameter are taken as values of the odds ratio. Thus, *Exp (β)* indicates how likely (or how many % more) the variable *Y* is observed under the influence of the variable X_P. The significance of the coefficient β_P is also considered as the significance of $OR_P = Exp(\beta_P)$ (Uzmay et al., 2001; Montgomery et al., 2012; Güneri and Durmuş, 2020).

Probit regression model

Probit analysis, as an alternative to logistic regression, is a model used to find the effect of one or more explanatory variables on a binary categorical response variable (alive-dead, working-not working, sold or not sold, etc.). Both logistic and probit regression analyses are quite similar to each other and the obtained probability estimates are close to each other. Log-odds are used in logistic regression analysis, while cumulative normal distribution is used in probit regression. In other words, the probit is the opposite of the cumulative standard normal distribution (Caffo and Griswold, 2006).

The assumption underlying the probit model is that the response function is in the form of $Y_i = \alpha + \beta X_i + U_i$. Here, X_i is a variable that can be observed but Y_i is a variable that cannot be observed. If *Yi > 0*, Y_i=1, but if $Y_i < 0$, $Y_i = 0$. When assigning the result of the Y_i variable, the c value used as the threshold value is usually taken as zero (0), but another number value can also be used instead of zero (Demaris, 2004). If the cumulative normal distribution function *Φ (z)* is defined as *Φ (z) = P (Z≤z)* for the normal standard variable Z, it can be expressed as:

$$P(Yi = 1) = P(U_i > -\alpha - \beta X_i) = 1 - \emptyset(\frac{-\alpha-\beta X_i}{\sigma}) \qquad (Eq.7)$$

$$P(Yi = 0) = P(U_i \leq -\propto -\beta X_i) = \emptyset(\frac{-\propto -\beta X_i}{\sigma}) \qquad \text{(Eq.8)}$$

When there is more than one explanatory variable in the Probit model, it is defined as $Pr(Y=1/X) = \emptyset(X\beta)$. Here, Φ is the standard normal probability distribution. βX is called Probit score or index and has a normal distribution. Probit coefficient (β) refers to an increase that a unit increase in the estimate will make in the Probit score. The Probit coefficient measures the effect that the independent variable will have on the standard z-value of the dependent variable. The numerical magnitudes of these coefficients have no importance and no special interpretation; they only determine the direction and degree of the relationship (MacKinnon et al., 2007; Kulendran and Wong, 2011).

Results and Discussion

Description of dependent and independent variables belonging to 180 cattle randomly determined from the enterprises and their descriptive statistics are given in *Table 3*. In the table, the distribution of each variable by breeds, their means, standard deviations, standard errors, and the confidence interval values (limits) are summarized.

Table 3. Descriptive statistics for dependent and independent variables

Variable	(X_6) *Breed	Total Cow	Cow with mastitis	Proportion	SD	95% CI for p Lower - Upper
(Y) Mastitis	1	41	23	0.561	-	0.407-0.715
	2	67	32	0.477	-	0.357-0.598
	3	72	30	0.417	-	0.302-0.531
(X_4) Infectious diseases	1	41	23	0.561	-	0.407-0.715
	2	67	34	0.508	-	0.387-0.628
	3	72	38	0.528	-	0.412-0.644
(X_5) Nutrition	1	41	17	0.415	-	0.262-0.567
	2	67	29	0.433	-	0.313-0.552
	3	72	38	0.528	-	0.412-0.644
				Mean	**SD**	**95% CI for μ**
(X_1) Live weight	1	41	-	465.76	31.83	456.2- 475.3
	2	67	-	580.90	34.09	573.4- 588.4
	3	72	-	603.96	27.14	596.8- 611.2
(X_2) Environmental hygiene	1	41	-	43.05	24.05	36.05- 50.05
	2	67	-	49.57	20.98	44.09- 55.04
	3	72	-	53.47	23.48	48.19- 58.76
(X_3) Milking hygiene	1	41	-	52.39	20.74	46.02- 58.76
	2	67	-	55.24	20.60	50.25- 60.22
	3	72	-	58.85	20.71	54.04- 63.66

(X_6) = * Breed was coded as "1, 2, and 3": 1. Local breed, Southern Anatolian Red (SAR), 2. Holstein Friesian and 3. Simmental, SD: standard deviation, CI: confidence interval

Data of the study were analyzed and evaluated by using the Stata (2018) package program. The data were analyzed according to both the probit and logistic models. The fit statistics of the statistical results of the probit model obtained from the definition

interval in the *Eq.7* and *Eq.8* are given in *Table 4*. Similarly, some fit statistics of the logit model are given in the same table for comparison.

Table 4. *Some goodness of fit statistics for Logistic and Probit regressions*

Goodness of fit criteria	Probit regression	P	Logit regression	P
LR CHI^2 (DF=7)	177.480	0.000	**179.770**	0.000
AIC	0.519	-	**0.507**	-
PSEUDO R^2	0.7128	-	**0.722**	-
Cragg & Uhler's R^2	0.837	-	**0.843**	-
Mckelvey and Zavoina's R^2	0.845	-	**0.859**	-
Log Likelihood	-35.749	-	**-34.605**	-
Hosmer Lemeshow CHI^2 (DF=8)	-	-	**12.212**	0.142

AIC: Akaike information criterion, LR chi^2: likelihood ratio chi-square test, DF: Degrees of freedom, P: probability

In general, the Probit and Logit regression analyses gave results close to each other. However, since we had to choose one of them in the interpretation of variables, it was required to look at the adaptation criteria. As can be seen in *Table 4*, all goodness of fit criteria and test criteria mainly point to the logit model. For example, the fact that the AIC value of the logit model is lower than the probit (0.519>0.507) indicates that logit is a better model. In addition, the fact that the Log Likelihood value of the logit model is higher than the probit model (-35.749<-34.605) again indicates that the logit model should be preferred. Moreover, the value of LR chi2 (179.77), which again plays an important role in the choice of the model, should be large and significant (p<0.01). Taking into account other criteria like these, it was decided to prefer the logit model and make evaluations on the results of this model.

The results obtained by running the following command in Stata (2018) program are given in *Table 5*. In addition, OR (Odds ratio Exp(B)) values obtained from *Eq.6* are given in the same *Table 5*:

$$\text{Logit } Y\ (X_1)\ (X_2)\ (X_3)\ i.(X_4)\ i.(X_5)\ i.(X_6)$$

Here, the infectious disease variable (X4) was significant at the level of p<0.05, while all other independent variables were significant at the level of p<0.01. This indicates that the independent variables are adequate in explaining the dependent variable.

Table 5. *Parameter estimates for the Logistic regression model*

Variables	Coefficient B	SEM	Wald	df	P	Odds ratio Exp(B)	95% CI for Exp(B) Lower	Upper
(X$_1$) Live weight	0.043	0.012	12.721	1	0.000	1.044	1.019	1.068
(X$_2$)Environmental hygiene	-0.064	0.016	16.292	1	0.000	0.938	0.909	0.968
(X$_3$) Milking hygiene	-0.120	0.023	27.952	1	0.000	0.887	0.848	0.927
(X$_4$) Infectious disease (1)	1.449	0.671	4.659	1	0.031	4.257	1.143	15.859
(X$_5$) Nutrition (1)	-2.975	0.710	17.550	1	0.000	0.051	0.013	0.205
(X$_6$) Breed			9.643	2	0.008			
(X$_6$) Holstein(2)	-4.927	1.622	9.226	1	0.002	0.007	0.000	0.174
(X$_6$) Simmental(3)	-5.632	1.855	9.215	1	0.002	0.004	0.000	0.136
Constant	-9.915	5.256	3.559	1	0.059	0.000		

SEM: standard error of means, df: degrees of freedom, P: probability, CI: confidence interval

Whether there is a specification error in the created logistics model can be understood by entering the 'linktest' command in the Stata (2018) program. The obtained results are given in *Table 6*, and the '*hatsq*' statistic was not found significant (p>0.05). This indicates that there is no specification error in the model. That is, there is no missing variable to be included in the model or an unnecessary variable to be removed from the model. Here, the variable '*hat*' shows the estimated values of the dependent variable. '*hatsq*', on the other hand, shows the variable $\widehat{Y}_i^2$, consisting of the square of estimated values (Mert, 2016).

Table 6. *Test results for the specification error*

Mastitis	Coefficient	Standard error	Z	P	[95% Conf. Interval]
hat	1.050	0.172	6.08	0.000	0.712 - 1.389
hatsq	0.051	0.046	1.11	**0.267**	- 0.039 - 0.142
constant	-0.219	0.379	-0.58	0.562	-0.963 - 0.523

Z: Z test statistic, P: probability

Before interpreting the logistic model, the '*collin*' command was used to check whether there was multicollinearity between continuous variables in the regression model. For this, the following command was run in the Stata (2018) and the results showing the vif values were obtained (*Table 7*):

collin X_1 X_2 X_3

Here, the fact that vif values for the live weight(X_1), milking hygiene(X_2), and environmental hygiene (X_3) are below five (vif<5) indicates that there is not any multicollinearity. Since there can be no multicollinearity between categorical variables, there is no need to calculate vif values.

Table 7. *Collinearity diagnostics*

Variable	VIF	SQRT VIF	Tolerance	R-Squared
(X_1) Live weight	1.02	1.01	0.9791	0.0209
(X_2) Envir. hygiene	1.07	1.04	0.9309	0.0691
(X_3) Milking hygiene	1.05	1.03	0.9498	0.0502
Mean VIF	1.05			

SQRT VIF: square root of the variance inflation factor

When the parameters obtained based on the logistic model are given in *Table 5*. It was observed that the coefficients of the live weight variable and the infectious disease variable were positive and the coefficients of the other independent variables were negative. In this case, looking at the odds ratio of the live weight variable, it can be interpreted that a one-unit increase in live weight increases the likelihood of mastitis by 1.044 times. In other words, the probability of getting mastitis increases by 4.4%. As the age of the animal's increases, also, there is a natural increase in body weight, milk yield and udder tissues. Due to the fact that the udder tissues, especially the nipples are close to the ground and contact surface with the bedding material and feces increases and this situation reveals the risk of mastitis (Kuczaj, 2003).

The coefficient of environmental hygiene is negative, and by looking at its odds ratio, it can be interpreted that a one-unit increase in environmental hygiene scores leads to a 6.2% (1-0.938) decrease in the likelihood of being mastitis. Similarly, it can be also interpreted that a one-unit increase in milking hygiene scores will reduce the likelihood of being mastitis by 11.3% (1-0.887). Ndahetuye et al. (2020) stated that paying attention to environment and milking hygiene may reduce the risk of mastitis in dairy cows with a high rate of subclinical mastitis. In addition, many researchers reported that the dryness of the bedding material that cattle are in constant contact with reduces the rate of contamination in the udder tissue. The reduction of mastitis pathogen risk in dry bedding material supports the results of the current study (Jobim et al., 2010; Lago et al., 2011; Oliveira and Ruegg, 2014; Fávero et al., 2015).

It has been reported that streptococci which cause infectious diseases, are one of the main causes of subclinical mastitis in cattle herds (Jobim et al., 2010; Oliveira and Ruegg, 2014). In this study, statistics of data (*Table 5*) show that cattle showing signs of infectious disease are 4.257 times more at risk of mastitis than healthy cattle. Alpay and Yeşilbağ (2009) and Zadoks et al. (2001) reported that there is a high degree of positive correlation between infectious diseases caused by many microorganisms and mastitis. In addition, these investigators stated that many secretions and spills such as nose, eyes, saliva, stool, semen and vaginal discharge and milk of infected animals, and placenta and fetal tissues play an important role in the spread of the virus to environment. This situation, explains why the risk of mastitis is high in cattle with symptoms of infectious disease.

Cattle fed adequately with a good and balanced ration are 94.9% (1 - 0.051) less likely to get mastitis compared to cattle fed inadequately and unbalanced (*Table 5*). Ingvartsen and Moyes (2013) stated that feeding with a balanced ration may prevent sudden changes in body condition score, so, it increases resistance against diseases by keeping immune cell functions at a high level. Inadequate nutrition in terms of vitamin E and selenium can lead to an increase in infections in cows (Cengiz, 2009). In addition, Spears and Weiss (2008) stated that the incidence of mastitis was reduced by 37% and clinical symptoms by 62%, by giving oral selenium and vitamin E (740 IU/day) for 3 weeks before delivery. This supports the finding of how effective nutrition is on mastitis.

In *Table 3*, mastitis rates of each breed are given. Here, the highest incidence of mastitis is in local breeds (56.1%). This is followed by Holstein (47.7%) and Simmental (41.7%), respectively. The reason for this is that farmers give more attention to care for cultural breeds, and it has been observed that they behave more carefully towards them. In this experiment, local breed (SAR) was accepted as a reference, however, it was statistically insignificant (p>0.05), so, local breed didn't compare with Holstein and Simmental breed's mastitis occurrence. However, the possibility of mastitis occurrence for each breed are given in *Table 9* and have been interpreted separately.

In *Table 8*, the classification of cows according to whether they have mastitis are given. The predictive power of the Logit model can be seen in this classification table. In the *Table 8*, out of a total of 95 cows without mastitis disease, 87 were classified correctly and 8 were classified incorrectly, which indicates that estimation is carried out at a 91.6% accuracy level. Similarly, out of the total 85 cows with mastitis disease, 80 were correctly classified and 5 were incorrectly classified, which shows that the correct estimate rate is 94.1%. Overall, it can be said that the model has a high estimation rate (92.8%).

Table 8. Classification table

		Mastitis		Percentage correct
		No	Yes	
Mastitis	No	87	8	91.6
	Yes	5	80	94.1
Overall percentage				92.8

To estimate the impact rates of each of the categorical variables (infectious disease (X_4), nutrition(X_5), and breed (X_6)) on mastitis disease, the following 'margins' command was used in the Stata (2018) program, and the results given in *Table 9* were obtained:

margins X_4 X_5 X_6, at means

It is seen in the *Table 9* that all factors except the local breed factor have a significant $(p<0.05)$ impact rate on mastitis.

Table 9. Risk probabilities of categorical independent variables that are effective in the mastitis disease

Categorical variable	Impact rate margin	Delta-method Std. Err.	Z	P	95% confidence interval	
(X_4) Infectious Dis.					Lower	Upper
0	0.217	0.089	2.42	0.015	0.041	0.393
1	0.541	0.107	5.03	0.000	0.330	0.753
(X_5) Nutrition						
0	0.705	0.097	7.26	0.000	0.514	0.895
1	0.108	0.050	2.15	0.032	0.009	0.208
(X_6) Breed						
(1) Local (SAR)	0.972	0.071	1.59	0.113	-0.026	0.252
(2) Holstein	0.204	0.091	2.24	0.025	0.025	0.383
(3) Simmental	0.112	0.036	26.82	0.000	0.901	1.043

Z: Z test statistic, P: probability

As can be seen in *Table 9*, the probability of getting mastitis in cows without any infectious disease (0) is 21.7%, while the probability of getting mastitis in cows with symptoms of any infectious disease increases to 54.1%. Since the body resistance of cattle fed inadequately with a poor and unbalanced ration (0) weakened, the probability of getting mastitis is estimated as 70.5%. However, in cattle fed balanced with a ration (1) whose nutrients, minerals, and vitamins balance is well-adjusted, the probability of getting mastitis decreases to 10.8%. These results are consistent with the study of Sundrum (2015) who stated that the incidence of mastitis increased up to 60% in herds with poor care, feeding and adverse environmental conditions throughout the world. On the other hand, in an experiment carried out in the same region (Tel et al., 2009), rate of mastitis was quite high (72.4%) compared to other parts of Turkey. The researchers claimed that this was due to the lack of environmental and milking hygiene and dry period treatment habits, especially in small scale family farms.

When the breed factor was considered, the probability of getting mastitis was 11.2% in the Simmental breed, while the rate increased to 20.4% in the Holstein breed (*Table 9*).

Koç (2016) states that, the somatic cell count of the Simmental breed in raw milk is 70 to 80% lower than the Holstein breed, and Simmental breeds are more resistant against to mastitis. Tel et al. (2009) reported that the incidence of mastitis in Simmental cows in Turkey was 15.78%. This data supports the research findings obtained from the region. In local breeds, this rate was found to be 97.2%, but it was not significant (p>0.05). When the reason of this was investigated, it was understood that farmers are take care more to cultural breeds that are more economically valuable, but they ignored many factors, such as environmental conditions, care, milking hygiene, and nutrition, for local breeds because they approach these cattle traditionally. This leads to a decrease in cows' milk yields by increasing the likelihood of developing subclinical mastitis that farmers cannot see by eye. Tel et al. (2009) recommended that for overcome of these problems the breeders in that region should be trained.

In order to estimate the impact rates of milking (X_3) and environmental hygiene (X_2) on mastitis, it is necessary to categorize the variables (*Table 10*). Since scores are not categorical variables, they must be categorized by classifying first (Schreiner and Ruegg, 2003; Mert, 2016). Then, it is possible to obtain the impact probability of the score value of each created class on mastitis by using the 'margins' command. For environmental hygiene (X_2) and milking hygiene (X_3), the following commands were run respectively in the Stata 14.2 program:

margins, at(X₂ =(20(20)100)) vsquish
margins, at(X₃ =(20(20)100)) vsquish

Here, the environment and milking hygiene, divided into 5 classes at 20 point intervals, were categorized, and the impact rates (margins) of each category on mastitis disease are obtained as in *Table 10*.

Table 10. *Impact rates of environmental and milking hygiene scores on mastitis*

Class scores	Impact rate margin	Delta-method Std. Err.	Z	P	95% confidence interval	
Environmental hygiene					Lower	Upper
1-(20%)	0.609	0.038	15.99	0.000	0.534	0.683
2-(40%)	0.518	0.026	19.89	0.000	0.467	0.569
3-(60%)	0.424	0.028	15.13	0.000	0.369	0.479
4-(80%)	0.324	0.046	7.00	0.000	0.233	0.415
5-(100%)	0.220	0.065	3.35	0.001	0.091	0.349
Milking hygiene						
1-(20%)	0.862	0.053	16.17	0.000	0.758	0.967
2-(40%)	0.643	0.043	14.67	0.000	0.557	0.729
3-(60%)	0.397	0.036	10.99	0.000	0.326	0.468
4-(80%)	0.179	0.057	3.12	0.000	0.066	0.291
5-(100%)	0.040	0.035	1.14	0.256	-0.029	0.111

Z: Z test statistic, P: probability

As seen in *Table 10*, the probability of getting mastitis is 60.9% in enterprises with the lowest environmental hygiene (20 points), while the risk of getting mastitis also decreases as scores increase. The probability of getting mastitis of cows in the enterprise that receives a 100 score from environmental hygiene falls to 22%. Similarly, while the probability of mastitis is 86.2% in enterprises that receive 20 points from milking hygiene,

the risk of mastitis decreases in enterprises that pay more attention to the milking procedure. That is, in an enterprise that scores 80 by paying more attention to milking hygiene conditions, the probability of mastitis falls to 17.9% (p<0.01).

Mastitis may develop in the udder lobes of dairy cows inadequate care before and after milking and non-hygienic conditions. While mastitis can develop due to a single factor, it is inevitable that the risk of mastitis increases with the combination of negative factors.

Teat dipping solution before and after milking are important not only to reduce the possibility of mastitis but also to reduce the risk of bacterial contamination in milk (Zucali et al., 2011). Before milking, the teats should be dipped in an antiseptic solution and waited for at least 30 seconds, then the teats should be wiped with disposable paper napkins or towels and dried. If not dried, the remnants of the solution applied to the udder contaminate the milk. If the above-mentioned precautions are not taken before milking, microorganisms pass into the milk. In this case, since the quality of milk deteriorates, it causes odor, taste, structure and color changes in products obtained from milk, causing a decrease in product quality (Nelson and Trout, 1964).

Since the teats remain open for a while after milking, flies and microorganisms easily transmitted by animal manure increase the risk of mastitis. To protect from this contamination, after milking, teat dipping should be done, and the cows should be fed so that they can stand for at least half an hour.

The estimated logit model of factors affecting the mastitis disease was defined as follows, and it was determined that the model had a high prediction rate of 92.8%.

Model for mastitis = 1

$$Z = -9.914 + 0.043 * (\text{LiveW.}) - 0.064 * (\text{EnvirH.}) - 0.12 * (\text{MilkingH.}) + 1.449 * (\text{Inf. Dis} = 1) - 2.975 * (\text{Nutrition} = 1) - 4.927 * (\text{Breed} = 2) - 5.631 * (\text{Breed} = 3)$$

It may be a little difficult for the model to be applied directly by the farmers in practice. However, it can be easily used by a veterinarian or zootechnist. The Z value is calculated by defining the relevant parameters in the *Eq.3* model, and by replacing them in *Eq.1*, estimation values in the range of $0 \le P(Y) \le 1$ could be obtained. The obtained value gives the probability of getting mastitis of the cow under the mentioned living conditions.

Conclusion and Recommendations

In conclusion, in this study, the impact rates of factors that directly and indirectly affect mastitis disease, which secretly gnaws on the udder lobes of dairy cows and causes a constant decrease in milk yield, was revealed. It was attempted to draw attention to the effects of these factors, which are ignored especially by small family enterprises, on mastitis. It should be told to enterprise managers that taking the necessary measures for the mentioned factors is a cheaper way than treating animals. In addition, it was concluded that although they were more adapted to environmental conditions, local breeds, which are fed poorly in pastures, are far from veterinary control and whose living and caring conditions were not paid attention, were at risk more in terms of mastitis.

In cases where the dependent variable is in a binary structure, the LS method is insufficient in estimating the relationships between the dependent and independent variables since it cannot provide the necessary assumptions. Therefore, it is necessary to resort to logistic regression analysis, which does not require the assumption of normality and continuity, and also offers a prediction performance as good as machine learning.

When the study is applied to larger integrated farms in the future, it can ensure that measures are taken to minimize economic losses. However, there are many other factors that directly and indirectly affect the risk of mastitis. For example, these include disease types, ration content, number of lactations, genetic predisposition, milking number, periodic mastitis controls, environmental factors such as milk yields, climatic conditions, temperature, humidity, and technical factors such as the breeder's expertise and educational status. For this reason, it would be beneficial to conduct more comprehensive studies covering all factors by researchers.

REFERENCES

[1]   Alpay, G., Yeşilbağ, K. (2009): The roles of viruses in mastitis. – Uludag Univ. J. Fac. Vet. Med. 28(1): 39-46.

[2]   Caffo, B., Griswold, M. (2006): A user-friendly introduction to link-probit-normal models. – The American Statistician 60(2): 139-145.

[3]   Cengiz, M. (2009): The protective approaches against dry period mastitis in cows. – Atatürk Univ. Journal of Veterinary Sciences 4(3): 215-222.

[4]   Chebel, R. C., Santos, J. E. P., Reynolds, J. P., Cerri, R. L. A., Juchem, S. O., Overton, M. (2004): Factors affecting conception rate after artificial insemination and embryonic loss in lactating dairy cows. – Anim. Reprod. Sci. 84: 239-255.

[5]   Cobirka, M., Tancin, V., Slama, P. (2020): Epidemiology and classification of mastitis. – Animals 10(12): 2212.

[6]   Demaris, A. (2004): Regression with Social Data: Modeling Continuous and Limited Response Variables. – John Wiley & Sons, Inc. Hoboken, New Jersey.

[7]   Fávero, S., Portilho, F. V. R., Oliveira, A. C. R., Langoni, H., Pantoja, J. C. F. (2015): Factors associated with mastitis epidemiologic indexes, animal hygiene, and bulk milk bacterial concentrations in dairy herds housed on compost bedding. – Livestock Science 181: 220-230.

[8]   Guimaraes, J. L., Brito, M. A., Lange, C. C., Silva, M. R., Ribeiro, J. B., Mendonça, L. C., Souza, G. N. (2017): Estimate of the economic impact of mastitis: A case study in a Holstein dairy herd under tropical conditions. – Prev Vet Med. 142: 46-50.

[9]   Gujarati, D. N., Porter, D. C. (2009): Basic Econometrics. – 5th Edition, McGraw Hill Inc., New York.

[10]   Güneri, Ö. İ., Durmuş, B. (2020): Dependent dummy variable models: An application of logit, probit and tobit models on survey data. – International Journal of Computational and Experimental Science and Engineering (IJCESEN) 6(1): 63-74.

[11]   İnal, M. E., Topuz, D., Uçan, O. (2006): Linear probability and parameter estimation with Logit models. – Sosyoekonomi-1 3(3): 47-72.

[12]   Ingvartsen, K. L., Moyes, K. (2013): Nutrition, immune function and health of dairy cattle. – Animal 7: 112-122.

[13]   Jamali, H., Barkema, H. W., Jacques, M., Lavallée-Bourget, E. M., Malouin, F., Saini, V., Stryhn, H., Dufour, S. (2018): Invited review: Incidence, risk factors, and effects of clinical mastitis recurrence in dairy cows. – Journal of Dairy Science 101(6): 4729-4746.

[14]   Jobim, M. B., Lopes, M. A., Costa, G. M. D., Demeu, F. A. (2010): Pathogens associated with bovine mastitis in dairy herds in the south region of Brazil. – Animal Industry Bulletin 67: 175-181.

[15]   Kılıç, B., Keskin, İ. (2019): Determination of factors effective in diagnosis of mastitis in holstein cattle by logistic regression analysis. – Journal of Bahri Dagdas Animal Research 8(2): 46-55.

[16]   Koç, A. (2016): A review on Simmental raising: 2. studies in Turkey. – Journal of Adnan Menderes University Agricultural Faculty 13(2): 103-112.

[17] Kuczaj, M. (2003): Analysis of changes in udder size of high-yielding cows in subsequent lactations with regard to mastitis. – Electronic J. Pol. Agric. Univ., Ser. Anim. Husb. 6(1): 2-5.

[18] Kulendran, N., Wong, K. K. F. (2011): Determinants versus composite leading ındicators in predicting turning points in growth cycle. – Journal of Travel Research 50(4): 417-430.

[19] Lago, A., Godden, S. M., Bey, R., Ruegg, P. L., Leslie, K. (2011): The selective treatment of clinical mastitis based on on-farm culture results: I. Effects on antibiotic use, milk withholding time, and short-term clinical and bacteriological outcomes. – J. Dairy Sci. 94: 4441-4456.

[20] Lakew, M., Tolosa, T., Tigre, W. (2009): Prevalence and major bacterial causes of bovine mastitis in Asella, South Eastern Ethiopia. – Tropical Animal Health and Production 41(7): 1525-1530.

[21] Long, J. S., Freese, J. (2014): Regression Models for Categorical Dependent Variables Using Stata. – Third Edition, College Station: Stata Press, USA.

[22] Longo, F., Salat, O., Van Gool, F. (2001): Incidence of clinical mastitis in French dairy herds, epidemiological data and economic costs. – Folia Veterinaria 45: 45-46.

[23] MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., Hoffman, J. M. (2007): The ıntermediate endpoint effect in logistic and probit regression. – Clinical Trials 4(5): 499-513.

[24] Mekonnen, S. A., Koop, G., Melkie, S. T., Getahun, C. D., Hogeveen, H., Lam, T. J. G. M. (2017): Prevalence of subclinical mastitis and associated risk factors at cow and herd level in dairy farms in North-West Ethiopia. – Prev Vet Med. 145: 23-31.

[25] Mert, M. (2016): SPSS-STATA Horizontal Cross-Section Data Analysis Computer Applications. – Detay Yayıncılık, Ankara.

[26] Montgomery, D. C., Peck, E. A., Vining, G. G. (2012): Introduction to Linear Regression Analysis. – 5th Edition, Wiley, New York.

[27] Ndahetuye, J. B., Twambazimana, J., Nyman, A. K., Karege, C., Tukei, M., Ongol, M. P., Persson, Y., Båge, R. (2020): A cross sectional study of prevalence and risk factors associated with subclinical mastitis and intramammary infections, in dairy herds linked to milk collection centers in Rwanda. – Prev. Vet. Med. 179: 1-9.

[28] Nelson, J. A., Trout, G. H. (1964): Judging Dairy Products. – The Olsen Publishing Co. Milwaukee 12, Wis. U.S.A.

[29] Nusinovici, S., Zhang, L., Chai, X., Zhou, L., Tham, Y. C., Vasseneix, C., Majithia, S., Sabanayagam, C., Wong, T. Y., Cheng, C. Y. (2020): Machine learning to determine relative contribution of modifiable and non-modifiable risk factors of major eye diseases. – British Journal of Ophthalmology, doi: 10.1136/bjophthalmol-2020-317454.

[30] Oliveira, L., Ruegg, P. L. (2014): Treatments of clinical mastitis occurring in cows on 51 large dairy herds in Wisconsin. – J. Dairy Sci. 97: 5426-5436.

[31] Özenç, E. (2019): Determination of risk factors associated with subclinical mastitis as detected by california mastitis test in smallholder dairy farms in Afyonkarahisar. – Kocatepe Veterinary Journal 12(3): 277-283.

[32] Özyurt, A. (2013): Effects on reproductive performance of mastitis cases in postpartum period in Holstein Friesian dairy cows. – Journal of Animal Production 54(1): 21-29.

[33] Rougoor, C. W., Hanekamp, W. J. A., Dijkhuizen, A. A., Nielen, M., Wilmink, J. B. M. (1999): Relationships between dairy cow mastitis and fertility management and farm performance. – Prev Vet Med. 39: 247-264.

[34] Sabuncuoğlu, N., Çolak, A., Akbulut, Ö., Tüzemen, N., Bayram, B. (2003): Relationships between CMT scores and some milk yield traits in Holstein Friesian and brown Swiss cows. – Atatürk University Journal of Agricultural Faculty 34(2): 139-143.

[35] Sabuncuoğlu, N., Çoban, Ö. (2006): Mastitis Economy. – Atatürk University J. Vet. Sci. 1(1): 1-5.

[36] Schreiner, D. A., Ruegg, P. L. (2003): Relationship between udder and leg hygiene scores and subclinical mastitis. – J. Dairy Sci. 86: 3460-3465.

[37] Spears, J. W., Weiss, W. P. (2008): Role of antioxidants and trace elements in health and immunity of transition dairy cows. – Vet J. 176: 70-76.

[38] Stata (2018): STATA 14.2. – StataCorp LLC 4905 Lakeway Drive College Station, TX 77845 USA.

[39] Sundrum, A. (2015): Metabolic disorders in the transition period indicate that the dairy cows' ability to adapt is overstressed. – Animals 5: 978-1020.

[40] Tel, O. Y., Keskin, O., Zonturlu, A. K., Arserim Kaya, N. B. (2009): Subclinical mastitis prevalance and determination of the antibiotics susceptibility in Şanliurfa region. – Fırat Univ. Veterinary Journal of Health Sciences 23(2): 101-106.

[41] Uzmay, C., Kaya, A., Kaya, İ., Akbaş, Y. (2001): Studies on prevalence of mastitis and factors affecting prevalence in herds of İzmir Holstein breeders association. 2. relationships between managerial practices and subclinical mastitis. – Journal of Agricultural Faculty of Ege Univ. 38(2-3): 71-78.

[42] Zadoks, R. N., Allore, H. G., Barkema, H. W., Sampimon, O. C., Wellenberg, G. J., Gröhn, Y. T., Schukken, Y. H. (2001): Cow and quarter level risk factors for Streptoccus uberis and Staphylococcus aereus mastitis. – J. Dairy Sci. 84: 2649-2663.

[43] Zucali, M., Bava, L., Tamburini, A., Brasca, M., Vanoni, L., Sandrucci, A. (2011): Effects of season, milking routine and cow cleanliness on bacterial and somatic cell counts of bulk tank milk. – Journal of Dairy Research 78(4): 436-441.