Turkish Journal of Agriculture - Food Science and Technology, 9(6): 1154-1158, 2021
DOI: https://doi.org/10.24925/turjaf.v9i6.1154-1158.4384

Turkish Journal of Agriculture - Food Science and Technology

Available online, ISSN: 2148-127X | www.agrifoodscience.com | Turkish Science and Technology Publishing (TURSTEP)

Determining the Factors Affecting 305-Day Milk Yield of Dairy Cows with Regression Tree

Serdar Genc[1,a,*], Mehmet Mendes[2,b]

[1]*Department of Agricultural Biotechnology, Faculty of Kırşehir Agriculture, Ahi Evran University, 40100, Kırşehir, Turkey*
[2]*Animal Science Department, Faculty of Agriculture, Çanakkale Onsekiz Mart University, 17100, Çanakkale, Turkey*
[*]*Corresponding author*

ARTICLE INFO

Research Article

Received : 19/04/2021
Accepted : 28/04/2021

Keywords:
305-day milk yield
Regression tree
Prediction
Dairy cows
Breed

ABSTRACT

The purpose of this study was to determine the factors affecting the 305-day milk yield of dairy cattle by using Regression Tree Analysis (RTA). The data set of this study consisted of 8 different cattle breeds grown in Turkey. Breed (B), Province (P), Lactation Length (LL), Service Period (SP), Dry Period (DP), Parity (PR), Calving Year (CY), Calving Age (CA) and Calving Month (CM) were used to predict the 305-day milk yield. Results of RTM showed that the usage of this method might be appropriate for determining the important factors that would be able to affect the 305-day milk yield (R^2=71.3%). It was seen that the most important factors affecting the 305-day milk yield were the Breed, Lactation Length, Province, and Parity. Therefore, those selected factors were more efficient than the others in predicting the 305-day milk yield. RTA results also indicated that the lowest milk yield was estimated for Jersey, Jersey Crossbred, and Yerli Kara. Among the highest 305-day milk yield cows, the milk yield estimates of the cows in the second, third, fourth, fifth, and the sixth parities were found significantly higher than that of the cows in the first and seventh parities.

[a] serdargenc1983@gmail.com https://orcid.org/0000-0002-1512-9072 | [b] mmendes@comu.edu.tr https://orcid.org/0000-0003-0109-9914

Introduction

Milk yield of dairy cattle, as in the other farm animals (i.e. sheep, goat, and water buffalo), may be affected by different genetic and environmental factors and the relations between those factors (Mendes and Akkartal, 2009). Milk yield is one of the majors concerns especially for the scientists in the field of animal breeding are focused on. Therefore, the researchers try to increase genetic progress by selecting higher milk yielding animals for the next generation (Berry et al., 2007; Mirtagioglu et al., 2008). In order to estimate genetic parameters, it is needed to get pedigree record of all cows. Since milk yield is also affected by different environmental factors such as lactation length, calving interval, service period, calving age, calving month, herds etc. these kind environmental factors should also be considered for selection programs along with genetic factors (Khalid et al., 2007; Kuthu et al., 2007). Therefore, determining the factors that will be able to affect the milk yield of dairy cattle is very important. There are different tests and approaches and mathematical models have been proposed for estimating milk yield of dairy cattle (Van Vleck and Henderson, 1961; Ashmawy et al., 1985). In practice, in many cases, various mathematical models are used by the researchers to estimate milk yield and genetic progress in the future lactations. However, the reliability of those mathematical models depends on many biological factors and thus those models will not be useful when these effects are not included in the model or not used correctly (Olori et al., 1999; De'ath and Katharina, 2000; David and Paul, 2004;Kocak et al., 2007; Zheng et al., 2009). However, different data mining techniques like Regression and Classification Tree, Artificial Neural Networks have been developed and these techniques may be effectively used in determining the factors that affect milk yield (Lacroix et al., 1995). In this study, it has been aimed at determining important factors that can affect the 305-day milk yield of different dairy cattle breeds by using Regression Tree Analysis Technique (RTA). Regression Tree Analysis (RTA) was used to determine the most important factors in predicting the 305-day milk yield of dairy cows (Mendeş, 2021).

Material and Methods

The data sets of this study were consisted of lactation records obtained from Cattle Breeder Association of Turkey. 9 different factors (Breed, Lactation Length, Service Period, Dry Period, Parity, Calving Year, Calving Age, Province, and Calving Month) of different dairy cattle were considered in investigating relations between 305 day milk yield and those factors that shown Table1.

Statistical Analysis

There are different techniques have been proposed in the literature for determining the factors affecting milk yield of farm animal, usage of the most appropriate method is extremely important in terms of reliability of the estimates. It is because, that way, it will be possible to get more detailed information about the effect of the factors and their interactions on the response. In the light of this, Regression Tree Method (RTM) has been used for determining the factor(s) that will be able to affect the 305-milk yield of different dairy cow breeds. RTA has been widely used for both prediction and classification in many fields of science such as medicine, industry, engineering, and agriculture (Mendeş and Akkartal, 2009; Çamdeviren et al., 2005; Karabağ et al., 2010). The use of this method in animal science is not common when compared to the other fields of science. However, the RTA, which has many advantages over the traditional methods, may be commonly and efficiently used in animal science studies. In this study we used RTA to determine the factors affecting 305-day milk yield of dairy cows and to predict milk yield by using some observed variables.

In this study, Regression Tree Analysis (RTA) was used to determine the most important factors in predicting the 305-day milk yield of dairy cows with SPSS package program (SPSS, 2008).

The purpose of RTA is to produce terminal nodes, which are homogeneous with respect to the target variable (Mendeş and Akkartal, 2009; David and Paul, 2004; Breiman et al., 1984; Bevilacqua et al., 2003; Çamdeviren et al., 2005; Karabağ et al., 2010). RTA finds the best possible variable or factor to split the node into two child nodes. CHAID algorithms were used. In choosing the best splitter, the program seeks to maximize the average ‘‘purity’’ of the two child nodes. More detailed information about the RTA can be found in Brieman et al., 1984.

Results and Discussion

The descriptive statistics of independent and dependent variables are given in Table 1. Figure 1 (optimal tree) shows the predictions of 305-day milk yield of dairy cows by using the factors including breed, lactation length, province, parity, service period, calving interval, dry period, calving age, calving year, and calving month. In Figure 1, node 0 is called the root node and it contains descriptive statistics related to 305-day milk yield. Firstly, the effect of each independent variable on the prediction of the 305-day milk yield was evaluated separately. For this, the importance of each variable was calculated. Because the Breed reflected the highest 305-day milk yield, it was determined to be the most important variable or factor, followed by Lactation length, Province and Parity. Therefore, among the 10 variables of factors, only 4 were selected. Using these 4 factors, we formed 6 terminal nodes. Each of these nodes was accepted as a homogenous group. Since Service Period, Dry Period, Calving, Calving Year, Calving Age, and Calving Month were not found to be effective in predicting the 305-day milk yield, these factors were not including to the optimal tree. Table 1 shows the risk value and its standard error. Risk value shows the variance within the nodes and it can be used as model fitness criterion. Therefore, the model which has a lower risk value will be a better model. The variance of the root node or dependent variable is $(1446.706)^2$ =2092958.25 and the risk value is 705312.43. In this case, the unexplained variation in the 305-day milk yield is found to be 0.337=33.7% (705312.43/2092958.25)=0.337. Therefore, the variation in the 305-day milk yield explained by the model will be 1-0.337=0.663=66.3% (Mendeş and Akkartal, 2009; Topal et al., 2010). It is concluded that 66.30% of the variation in 305-day milk yield can be explained by the four factors namely Breed, Lactation Length, Province, and Parity. As seen in the optimal tree (Figure 1), firstly, breeds in Node 0 or root node were divided into two nodes, based on Breed as Node 1 (Swedish Red, Montbeliarde-crossbreed) and Node 2 (Jersey, Jersey Crossbreed, Yerli Kara). As a result, Breed was the most effective factor in predicting the 305-day milk yield. The mean 305-day milk yield of the cows in Node 1 and Node 2 were predicted as 6133.341±23.955 and 2103.211±22.875, respectively. The proportions of the cows in Node 1 and Node 2 in total are 76.3% and 23.7%, respectively. Mean of the 305-day milk yield for cows in Node 1 were obviously higher than that of the Node 1. It is not sufficient, however, to use only Breed to predict the 305-day milk yield of the cows. In other words, cows in Node 1 was not homogeneous enough. Therefore, Node 1 generated by Breed in the first step was divided into nodes again based on Lactation length. Therefore, the Lactation length is accepted as the second most important factor in the prediction of 305-day milk yield of the cows.

Based on Lactation length values, 2958 cows in Node 1 were divided into two new nodes: Node 3 (≤293.5 day) and Node 4 (>292.5 day). As it can be seen from the optimal tree, the third, the fourth and the fifth important factors in predicting 305-day milk yield are Province, Breed, and Parity.

The mean of 305-day milk yield of the cows in the Provinces of 55 (Samsun), 6 (Ankara), and 15 (Burdur) is predicted as 5925.693 kg. In order to make reliable predictions for the cows in the Provinces of 35, 9, 10, 59, and 3 it is need to consider the Breed and Parity of the cows as well. As it is seen from the Node 10, the 305-day milk yield of the Montbeliarde-Crosbred cows is predicted as 5824.5 kg. For predicting the milk yield of the other breeds, on the other hand, the Parity also should be considered. When Node 11 and Node 12 are examined, it is seen that the cows the 305-day milk yield of the cows with the Parity of 2, 3, 4, 5, and 6 is significantly higher than that of the cows with the Parity of 1 and 7.

Results of this study showed that the most important factors affecting the 305-day milk yield were the Breed, Lactation Length, Province, and Parity.

Table 1. Descriptive statistics for independent variables

	N	305-day milk yield		Lactation Length		Service Period		Dry Period		Calving Age	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Breed											
Swedish Red	36	6936	181.0	361	15.3	124	15.6	43	3.1	35	1.8
Jersey	824	4320	28.8	328	2.2	110	2.3	63	0.8	54	0.9
Jersey cross.	51	4091	217.0	334	9.3	128	12.6	74	9.2	60	3.5
Red Holstein	823	6199	49.8	350	2.6	136	2.8	66	1.4	48	0.8
Red Holstein –Holstein cross.	341	6194	81.6	345	3.9	130	4.2	66	2.7	41	0.8
Montbeliarde	1448	6129	31.9	336	1.9	119	1.9	62	0.8	49	0.6
Montbeliarde cross.	265	5775	61.2	333	3.6	114	4.0	61	2.34	44	1.1
Yerli Kara	20	3924	323.0	353	21.2	139	24.8	67	16.2	53	5.1
Province											
Afyon	27	6429	264.0	300	8.1	85	8.8	66	3.6	52	5.2
Ankara	39	5576	276.0	336	10.7	118	11.0	62	1.7	45	2.5
Aydin	1918	6167	28.6	343	1.7	123	1.7	61	0.7	49	0.5
Balıkesir	106	5840	195.0	339	7.3	123	7.5	64	3.8	48	2.1
Burdur	384	5813	56.5	330	3.3	127	4.0	77	2.9	46	1.1
İzmir	316	6379	82.5	351	4.3	130	4.6	59	2.5	42	1.0
Samsun	872	4288	29.6	329	2.1	111	2.3	63	0.8	54	0.8
Tekirdag	146	6091	105.0	341	5.5	136	5.9	75	4.3	44	1.6
Parity											
1^{st}	1374	5591	37.9	344	2.0	127	2.1	64	1.0	29	0.1
2^{nd}	1041	5789	44.6	337	2.1	121	2.3	64	1.3	44	0.2
3^{rd}	650	5803	59.8	335	2.6	118	2.8	64	1.4	58	0.4
4^{th}	398	5739	72.3	331	3.3	113	3.4	62	1.4	72	0.5
5^{th}	154	5729	124.0	334	5.3	117	5.2	63	1.6	91	0.9
6^{th}	122	5738	134.0	343	7.1	123	6.8	61	1.6	98	1.3
7^{th}	69	5333	160.0	329	7.5	111	8.0	61	2.8	111	1.6
Calving Year											
2005	196	5798	102.0	321	4.3	115	5.2	75	4.0	47	1.5
2006	330	5850	81.9	339	3.7	127	4.0	67	2.0	45	1.1
2007	572	5668	60.1	342	3.1	129	3.3	68	1.6	49	1.0
2008	715	5702	55.6	340	2.7	124	2.8	64	1.3	49	0.9
2009	809	5665	53.3	345	2.6	129	2.8	64	1.4	49	0.8
2010	816	5760	48.3	341	2.4	120	2.5	59	1.1	50	0.8
2011	370	5530	69.9	316	2.6	92	2.5	56	0.9	50	1.1
Calving Month											
January	319	5866	81.8	345	4.2	132	4.5	67	2.4	48	1.1
February	383	5626	73.8	340	4.0	121	4.1	61	1.4	47	1.1
March	355	5501	78.7	340	3.8	121	3.9	62	1.4	51	1.3
April	320	5703	74.7	340	4.0	125	4.3	65	2.4	46	1.1
May	317	5536	76.0	339	3.8	126	4.2	67	2.3	45	1.2
June	327	5748	91.1	341	3.6	124	3.7	62	1.3	52	1.3
July	328	5707	77.4	333	3.6	114	3.6	61	1.6	50	1.3
August	286	5647	81.3	330	3.9	112	4.1	63	2.0	52	1.3
September	305	5787	83.4	331	3.8	115	4.1	64	2.3	51	1.3
October	316	5809	83.2	331	3.8	115	4.0	64	1.8	48	1.2
November	249	5719	92.6	339	3.9	120	4.1	62	1.7	50	1.6
December	303	5825	81.5	348	4.7	134	5.1	66	2.5	49	1.3
Overall	3808	5703	23.5	338	1.1	122	1.2	63	0.6	49	0.4

Table 2. Gain Summary for Nodes for Dependent Variable: 305 Daily MilkYield

Node	Node-by-Node			Cumulative		
	N	Percent	Mean	N	Percent	Mean
12	1003	25.9%	6539.61	1003	25.9%	6539.61
11	663	17.1%	6227.58	1666	43.0%	6415.43
7	303	7.8%	5925.69	1969	50.8%	6340.07
6	560	14.4%	5863.01	2529	65.2%	6234.43
10	190	4.9%	5824.50	2719	70.1%	6205.79
5	239	6.2%	5309.08	2958	76.3%	6133.34
2	920	23.7%	4299.30	3878	100.0%	5698.24

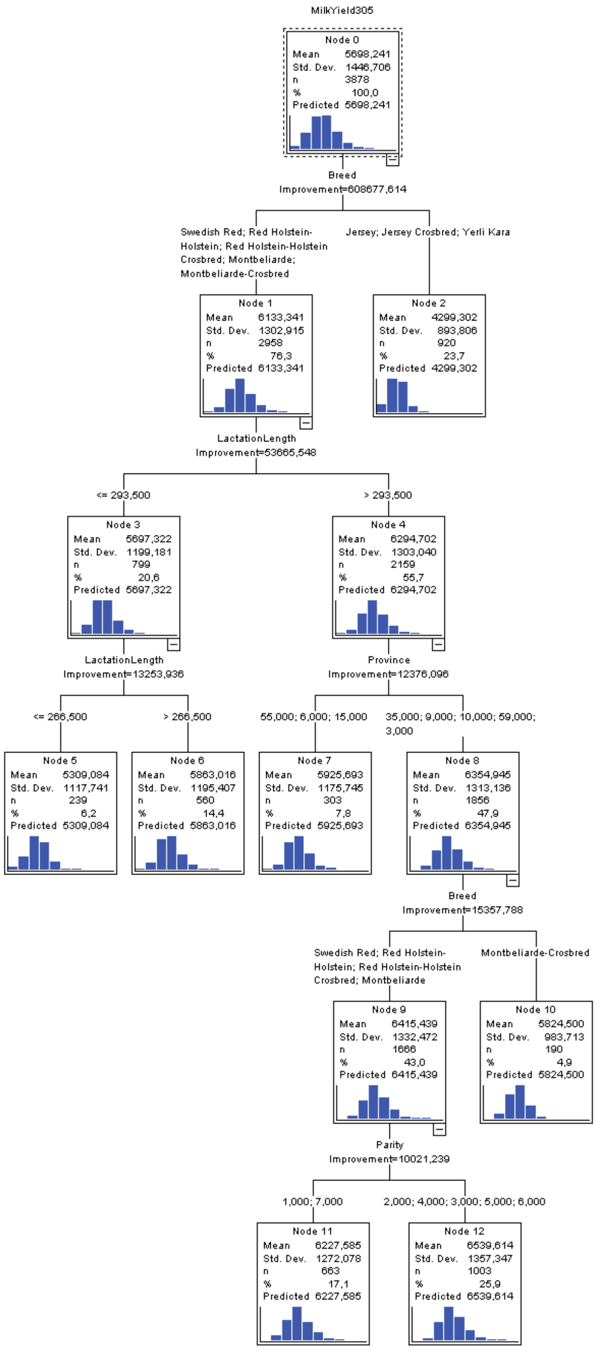

Figure 1. Structure of the Optimal Tree

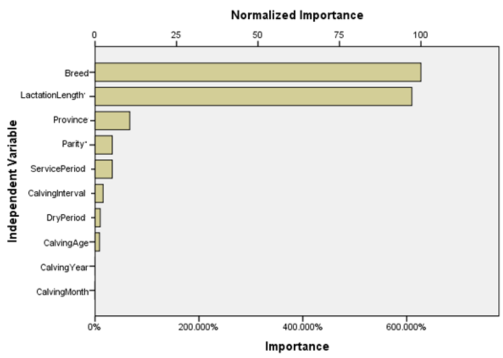

Figure 2. Normalized Importance for Independent Variables

Therefore, those selected factors were more efficient than the others in predicting the 305-day milk yield. It is thought that the effect of above factors on 305-day milk yield may change based on herd management, breeding systems, and maintenance and feeding. Furthermore, it is thought that the observed variation for the lactation length can be brought closer to the normal acceptable length (305 days) by arrangements to be made in production and marketing (Genç and Soysal 2018). Observed differences in the milk yield of the dairy cattle according to the provinces may have been due to the size of the farm, the maintenance and feeding conditions in the farms, the environmental conditions.

Although different results have been reported by several studies in terms of the effect of service period, type of birth, year of calving, age of calving and effect of birth season on 305-day milk yield, the effect of those factors were not found as significant in this study (Ulutas, 2002; Sahin et al., 2014; Genc and Soysal, 2019). This is due to the fact that the difference in the number of breeds. Only a single breed was generally considered by the previous studies, this study was carried out a very large data set with many breeds of dairy cattles in Turkey (Ulutas, 2002; Soydan and Sahin, 2016; Genc and Soysal, 2019).

Conclusion

One of the other important factors that caused to get different results was the differences in the statistical techniques which are used in analyzing data sets.In this study, the Regression Tree Method was used in determining important factors on 305-day milk yield. That way, it was possible to investigate the effect of latent and interrelated factors on milk yield estimation.

References

Ashmawy AA, Khattab AS, Hamed MK. 1985. Ratio and regression factors for predicting 305-day production from part-lactation milk records in Friesian cattle in Egypt. Bulletin of Faculty of Agriculture, Cairo University, Egypt.

Berry DP, Buckley F, Dillon P. 2007. Body condition score and live-weight effects on milk production in Irish Holstein-Friesian dairy cows. Animal, 1(9), 1351–1359.

Bevilacqua M, Braglia M, Montanari R. 2003. The classification and regression tree approach to pump failure rate analysis. Reliab Eng. Syst. Safe. 79, 59-67.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). Classification and regression trees. Chapman and Hall, Wadsworth Inc., New York, NY, USA.

Çamdeviren H, Mendeş M, Özkan, MM, Toros F, Şaşmaz T, Öner S. 2005. Determination of depression risk factors in children and adolescents by regression tree methodology. Acta Med. Okayama, 59(1), 19-26.

David RL, Paul LS. 2004. Multivariate regression trees for analysis of abundance data. Biometrics, 60, 543–549.

De'ath G, Katharina EF. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81(11), 3178–3192.

Genc S, Soysal MI. 2018. Milk yield and reproductive traits of Holstein cattle population in Turkey. Journal of Tekirdag Agriculture Faculty, 15 (1), 76-85.

Genc S, Soysal MI. 2019. Estimation of genetic parameters and genetic trend of Holstein Friesian cattle population in Turkey. Fresenius Environmental Bulletin, 28 (4), 2617-2624.

Karabağ K, Mendes M, Alkan S, Balcioglu MS. 2010. An assessment of embryonic mortality stages in Chukar partridge (Alectorischukar) by means of classification tree method. Archiv Fur Geflugelkunde, 74, 269-273.

Khalid J, Masroor EB, Muhammad A. 2007. Within herd phenotypic and genetic trend lines for milk yield in Holstein Friesian dairy cows. J. Anim. Biol., 1, 66-70.

Kocak S, Tekerli M, Ozbeyaz C, Yuceer B. 2007. Environmental and genetic effects on birth weight and survival rate in Holstein calves. Turk. J. Vet. Anim. Sci., 31(4), 241-246.

Kuthu ZH, Javed K, Ahmad N. 2007. Reproductive performance of indigenous cows of Azad Kashmir. J. Anim. Plants Sci., 17, 47-51.

Lacroix R, Wade KM, Kok R, Hayes JF. 1995. Prediction of cow performance with a connectionist model. Transactions of the American Society of Agricultural Engineers, 38 (5), 1573–1579.

Mendeş M. (2021). Re-evaluating the Monte Carlo Simulation Results by Using Graphical Techniques. Turkiye Klinikleri J Biostat. 13(1) doi: 10.5336/biostatic.2020-78896 (in pressing).

Mendes M, Akkartal E. 2009. Regression tree analysis for predicting slaughter weight in broilers. Italian Journal of Animal Science, 8 (4), 615-624.

Mirtagioglu H, Keskin S, Bakir G. 2008. Regression tree analysis for 305 day milk yield in Holstein cows. Indian Vet. J., 85, 943-945.

Olori VE, Hill WG, McGuirk BJ, Brotherstone S. 1999. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. Livest. Prod. Sci., 61, 53-63.

Sahin A, Ulutas Z, Adkinson YA, Adkinson RW. 2014. Genetic parameters of first lactation milk yield and fertility traits in Brown Swiss Cattle. Annals of Animal Science, 14(3), 545-557.

Soydan E, Sahin A. 2016. Estimates of genetic parameters for direct and maternal effects with six different models on birth weight of brown Swiss calves. Journal of Animal and Plant Sciences, 26 (3), 577-582.

SPSS, 2008. Statistical Package Social Science: SPSS for windows release 17.0, SPSS Inc., 2008.

Topal M, Aksakal V, BayramB, Yağanoğlu AM. 2010. An analysis of the factors affecting birth weight and actual milk yield in swedish red cattle using regression tree analysis. The Journal of Animal and Plant Sciences, 20 (2), 63-69.

Ulutas Z. 2002. Estimation of genetic and phenotypic trends of 305-day milk yield for Holsteins reared at Gelemen State Farm in Turkey. Indian Journal of Animal Sciences, 72 (10), 875-877.

Van Vleck, L.D., Henderson, C.R. (1961). Estimates of genetic parameters of some functions of part lactation milk records. Journal of Dairy Science, 44, 1073-1084.

Zheng H, Chen L, Han X, Zhao X, Ma Y. 2009. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. Agriculture, Ecosystems and Environment, 132, 98–105.