The Application of Ordinal Logistic Regression Model as a Robust Tool for Enhanced Prediction of Milk Yield in Dairy Cows

Sherif A. Moawed[1] , Esraa Mahrous[1*], Ahmed Elaswad[2], Hagar F. Gouda[3], and Ahmed Fathy[1]

[1]*Department of Animal Wealth Development, Biostatistics Division, Faculty of Veterinary Medicine, Suez Canal University, Ismailia 41522, Egypt*

[2] *Department of Animal Wealth Development, Genetics and Genetic Engineering Division, Faculty of Veterinary Medicine, Suez Canal University, Ismailia 41522, Egypt*

[3] *Department of Animal Wealth Development, Faculty of Veterinary Medicine ,Zagazig University, Zagazig 44511, Egypt*

* Corresponding author: esraa_mahrous@vet.suez.edu.eg +20 01008657263

Abstract:

Milk yield is a vital issue of concern in dairy cows. Hence, accurate milk production prediction is critical for improving dairy farm management and profitability. The purpose of this study was to examine the feasibility of applying ordinal logistic regression (OLR) to classify and predict milk production in Friesian cows into low (4500 kg), moderate (4500-7500 kg), and high (>7500 kg) classes. The data includes 3793 lactation records from dairy cows calved between 2009 and 2020 in order to investigate a number of explanatory variables, including the 305-day milk yield (305-MY), age at first calving (AFC), calving interval (CI), calving season (CFS), days open (DO), days in milk (DIM), dry period (DP), lactation order (LO), and number of services per conception (SPC). Significant determinants impacting yield were found, with varying impacts across different yield classes. The results suggested that LO, DIM, and 305-MY were the most significant parameters ($P < 0.05$) influencing data categorization. The OLR model demonstrated satisfactory fit in predicting milk yield categories, as it showed considerable accuracy (56%) and an area under curve equal to 0.69. In conclusion, the ordinal logistic regression demonstrated to be an effective method for modeling milk production as an ordinal parameter. The model's results provide insights into the complex interaction of factors influencing milk output, directing management strategies for optimal production.

Keywords: Ordinal logistic regression; Odds ratio; Dairy cows; Prediction; Milk production

Introduction

Milk production is the most attractive trait of dairy cows, and it keeps attracting the interest of researchers worldwide **(Garamu, 2019)**. According to **FAO (2023)**, Egypt produced about 5.2 million tons of milk in 2022. Friesian cows were renowned for their exceptional milk yield ability and form the foundation of many dairy farms. However, optimizing their productivity requires a thorough understanding of the complex interaction of factors impacting milk yield. Genetics, as well as environmental variables such as age, breed, season, lactation duration, calving interval, parity, stage of lactation, and days open, all impact dairy cow milk yield **(Susanto *et al.*, 2019)**. Hence, predicting and assessing milk production is crucial in the dairy sector for breeding and herd management evaluations to make strategic choices. Many mathematical methods were widely used to anticipate dairy milk production. However, due to the expanding database of animal records and advances in farm management software, new approaches were developed to handle more complex forms and include more input factors **(Zhang *et al.*, 2018)**.

There have been a lot of studies in the health and medical sciences over the last 30 years that commonly used ordered categorical data **(Ursino, 2014)**. The best methods for examining ordinal data are those that make complete use of the ordered Y-response; these are referred to as "Ordinal Regression Models" **(Akkuş *et al.*, 2019)**.

The use of ordinal logistic regression models (OLR) for predicting and estimating milk yield in dairy cattle is uncommon, as most studies attempt to predict exact milk production, whereas OLR divides the milk yield into categories **(Topuz, 2021)**. Therefore, we expected that using OLR to examine factors impacting milk yield in dairy cows, alongside odds ratio calculations and interpretations, might serve as a helpful guide for statistical modeling. This adaptable model may evolve extra factors, strengthening dairy producers and resulting in better practices and yields within the dairy industry.

Many studies have applied OLR in animal science. In the discipline of fertility, **Piles *et al.* (2013)** assessed the usefulness of OLR in enhancing the categorization of rabbit ejaculate in artificial insemination facilities, while **Peng (2019)** used the OLR technique to ascertain the influence of certain factors on abnormal sperm rate in boars. In the area of body condition score assessment, the goal of **Shittu *et al.***

(2014) was to assign body condition scores to the slaughtered animals based on a variety of criteria. Another trial used the morphometric traits to estimate the likelihood that adult equines will be overweight **(Martinson *et al.*, 2014)**. The application of OLR was extended to investigate the variables influencing farmers' adoption decisions. For example, to investigate the important aspects for adopting innovation in beef cattle production **(Abdullah *et al.*, 2021)**. Among clinical and epidemiological studies, **Ozawa *et al.* (2019)** used OLR to detect the physical factors associated with canine cognitive dysfunction.

As far as we know, few studies used OLR to predict and classify the milk production of dairy cows. Most studies on milk yield in dairy cows failed to recognize the ordered nature of milk production traits as a categorical variable. For example, **Grzesiak *et al.* (2003)** proposed a linear regression model that used test-day data, month of calving, and the proportion of Holstein-Friesian genes as characteristics to predict 305-days milk yield in addition to days in milk. Although, **Macciotta *et al.* (2002)**, **Vasconcelos *et al.* (2004), Valchev *et al.* (2020)**, and **Temesgen *et al.* (2022)** introduced regression-based models to estimate milk yield, they did not consider the ordered structure of milk yield features.

The OLR can be assessed using overall classification accuracy (ACC), which is the ratio of properly classified observations to total data **(Sokolova and Lapalme, 2009)**. Researchers frequently use additional metrics, such as sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), F1 score, Area under curve (AUC) to assess the predictive accuracy of classifiers **(Weng *et al.*, 2017, Kourou *et al.*, 2015)**. The range of these measurements were 0 to 1.

Therefore, the present study aimed to forecast Friesian cows' milk production by applying OLR and to assess the effectiveness of the OLR model in predicting and classifying milk yield classes.

Materials and methods

1. Data collection and Mmanagement:

In this study, 3793 dairy records were collected to explore the important variables impacting milk production in Holstein- Friesian cows that calved between 2009 and 2020. The data were obtained from one of Egypt's major dairy farms in El-Dakahlia province. Under an automated milking system, animals were mechanically milked three times each day. The total of these three numbers was used to calculate the total daily milk production. Cows with at least one parity were included in this investigation. Lactation records and all other data were computer-based and routinely updated. Healthy cows with normal births, no reproductive disorders, and no postpartum problems were used as study materials. All cows were periodically housed on clean floors, in open yards, or in slenderly covered yards that had access for cooling sprays.

2. Investigated characteristics:

The data of study were used to create OLR to identify the main factors influencing Friesians' milk production. The outcome variable of interest was therefore the total milk yield (kg), which was classified into three distinct groups: low (<4500 kg), moderate (4500–7500 kg), and high yield (>7500 kg). The rest variables were taken into consideration as independent variables, and these included the 305-day milk yield (305-MY/kg), age at first calving (AFC/month), calving interval (CI/month), calving season (CFS/spring–summer–fall–winter), days open (DO/days), days in milk (DIM/days), dry period (DP/days), lactation order of eight parities (LO/number), and number of services per conception (SPC/number). In addition, the previous authors' suggestions have been considered in the classification of the milk yield parameter and the selection of predictors **(Akkus and Ozkoc, 2012, Akkuş *et al.*, 2019, Akkus and Sevinc, 2020, El-Kasrawy *et al.*, 2020, Fathy *et al.*, 2023a, Moawed and Abd El-Aziz, 2022)**. Moreover, lower categories of categorical explanatory variables were suggested as baselines for other categories to better understand and explain variations in actual milk production.

3. Data preprocessing:

First, outliers and missing values were checked for the data. Three approaches can be applied to address outliers in a dataset: deleting them to trim the set, replacing or lowering their effect through outlier weight modifications, and applying robust procedures that are less susceptible to outliers such as OLR **(Kwak and Kim, 2017)**. The following procedures are used to deal with missing data: ignoring cases with missing data, feature selection with the elimination of variables that include missing values, and imputation of missing values, which is the most effective way **(Sessa and Syed, 2016)**. Second, the chi-square test was performed to pick the best factors for the multivariate analysis, with all parameters demonstrating high levels of significance ($p < 0.001$). Third, the data were randomly divided into two sets: the training set, comprising 80% of the data, which was used to create the model, and the testing set, comprising 20% of the data, which was used to validate and assess the performance of the model.

4. Assumptions:

The most important assumptions considered for applying the OLR include: (a) The dependent variable is measured on an ordinal scale. (b) The independent variables could be continuous, categorical, or ordinal.

(c) There is no multicollinearity. The variance inflation factor (VIF) was used to test the multicollinearity of suggested predictors; a VIF score of 1 implies no correlation, a VIF value between 1 and 5 suggests intermediate correlation, and a VIF value more than 5 indicates significant correlation **(Suleiman and Badamsi, 2019)**. (d) Proportional odds assumption (POA) exists, which indicates that each predictor has the same effect on the dependent factor. The brant test was used to verify this POA. The odds ratio (OR) is derived from dividing the probability of success (P) by the probability of failure (1-P). In practice, the OR is denoted by $exp\,(\beta)$ or (e^{β}), which refers for the exponent of each independent variable's coefficient. From the technical level, the independent variables in the datasets were divided into subcategories known as categories. With an OR of one, a certain category was selected as a baseline level. As a result, for each independent variable, the reference group was used as a benchmark for comparison with other categories of that variable. A positive association is shown by an estimated OR greater than one, and a negative relationship is indicated by an OR less than one **(Moawed and Abd El-Aziz, 2022)**.

5. OLR:

One of the most prevalent errors in statistical investigations is to construct a model without considering the data structure, statistical assumptions, and variable type. When a dependent variable in a regression has more than two categories and is measured in an ordered manner, OLR is used **(Malmquist and Rykatkin, 2023)**. In this study, an OLR model was used to classify milk production into high, moderate, and low categories and create a robust model for future prediction. The latent variable (Y*) approach is used for building OLR. Latent variables are attributes that cannot be directly examined but must be deduced from known covariations among a set of variables **(Tabachnick and Fidell, 2001)**. **Greene (2000)** defined the latent variable as the linear combination of the predictors.

$$Y^* = \sum_{k=1}^{K} b_\kappa X_\kappa + \varepsilon \quad \textbf{(1)}$$

In **Equation (1)**, b_κ indicates the estimated coefficient of the explanatory variable X_κ (305-MY, AFC, CS, CI, DIM, DO, DP, LO, SPC) and ε represents the model's error term. The milk yield variable in OLR is divided into J ordered categories, which are defined in respect to the latent variables and threshold parameters (μ), where "i" represents the model's ith observation **(Equation (2)**.

$$\begin{array}{ll} Y_i = 1, & Y^* \leq \mu_1 (= 0) \\ Y_i = 2, & \mu_1 < \ Y^* \leq \mu_2 \\ Y_i = 3, & \mu_2 < \ Y^* \leq \mu_3 \\ \ \ \vdots & \qquad \vdots \\ Y_i = J, & \mu_{J-1} < \ Y^* \end{array} \quad \textbf{(2)}$$

The threshold parameters are critical for identifying unknown ordered categories. Thus, the statistical significance of the threshold parameters, which was originally evaluated at the beginning of the analysis, ensures the ordinality of the dependent variables. If the actual model is an OLR, the computation of odds ratios, which can only be comprehended in logistic regression models, is also possible. The following equation expresses the chance that the milk yield variable will fall into category "j" given the predictors are present.

$$P(Y_i = j|x_{ik}) = F\left(\mu_j - \sum_{k=1}^{K} b_\kappa X_{i\kappa}\right) - F\left(\mu_{j-1} \sum_{k=1}^{K} b_\kappa X_{i\kappa}\right) \quad \textbf{(3)}$$

In **Equation (3)**, F represents the expected distribution function for the error term. To guarantee that all predicted probabilities were positive, a constraint $\mu_1 < \mu_2 < \mu_3 < \ldots \leq \mu_{J-1}$ must be applied. **Greene (2000)** suggested that the first threshold parameter (μ_1)) be set to "0". The following formula **(Equation (4))** can be used to assess the probability that the dependent variable belongs to category j or a lower category.

$$P(Y_i \leq j) = P(Y^* \leq \mu_j) = \frac{e^{(\mu_j - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}}{1+e^{(\mu_j - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}} \quad \textbf{(4)}$$

The OLR is obtained if the logistic distribution is represented by F in **Equation (3)**. The probability that the dependent variable falls into the appropriate groups is provided by **Equations (5), (6),** and **(7)**.

$$P(Y_i = 1) = \psi\left(-\sum_{k=1}^{K} b_\kappa X_{i\kappa}\right) = \frac{exp(-\sum_{k=1}^{K} b_\kappa X_{i\kappa})}{1+exp(-\sum_{k=1}^{K} b_\kappa X_{i\kappa})} \quad \textbf{(5)}$$

$$P(Y_i = 2) = \psi\left(\mu_2 - \sum_{k=1}^{K} b_\kappa X_{i\kappa}\right) - \psi\left(-\sum_{k=1}^{K} b_\kappa X_{i\kappa}\right)$$
$$= \left[\frac{exp(\mu_2 - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}{1+exp(\mu_2 - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}\right] - \left[\frac{exp(-\sum_{k=1}^{K} b_\kappa X_{i\kappa})}{1+exp(-\sum_{k=1}^{K} b_\kappa X_{i\kappa})}\right] \quad \textbf{(6)}$$

$$P(Y_i = J) = 1 - \psi\left(\mu_{J-1} - \sum_{k=1}^{K} b_\kappa X_{i\kappa}\right) = 1 - \left[\frac{exp(\mu_{J-1} - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}{1+exp(\mu_{J-1} - \sum_{k=1}^{K} b_\kappa X_{i\kappa})}\right] \quad \textbf{(7)}$$

The model's coefficients are calculated using maximum-likelihood principles, with the goal of finding coefficients that suit the data classification **(Minka, 2001).** All analyses were carried out using the R programming language and SPSS version 25.

Results

1. Tests of assumptions:

First, the study validated OLR's assumptions. The data contained no missing values. The box plot revealed that the data contained univariate outliers in most of the explanatory factors. A considerable level of multicollinearity was found using the VIF test, with VIF values between 2 and 5 for three parameters (DO = 5.84, DIM = 3.95, and CI = 2.36). A correlation analysis was done to find out which parameters were connected since VIF did not show which pair of predictors were correlated. The findings showed that DO was substantially associated with the CI (r=0.76) and had a significant association with DIM (r=0.85). While, the Brant test demonstrated that the POA is met. The null hypothesis—which holds that the coefficients of explanatory variables are constant across all milk yield categories—was accepted as the omnibus test revealed a statistically non-significant result ($P > 0.05$).

2. Ordinal logistic regression:

The multivariable OLR model results were displayed in **Table (1)**, which includes the estimated coefficients of the threshold and location parameters, standard errors, and OR. The actual milk production as a response variable is assumed to be an ordinal variable by the threshold parameters, which were statistically significant ($P < 0.001$). **Table (1)** shows that milk production was considerably impacted by lactation order ($P < 0.05$). In this study, all parities from the 2^{nd} to the 8^{th} parity except the 7^{th} parity had a negative OLR coefficients, meaning that the log odds of the response variable decreased by β units for each unit increase in the predictor variable. The level of milk yield is positively predicted by the 7^{th} parity. There is a predicted increase (0.702) in log odds in favor of a lower level of milk yield for every unit rise in 7^{th} parity. The OR for 7^{th} parity was 2.017, which is more than 1.

The results of calving seasons revealed that the milk yield increased significantly ($P<0.001$) in the winter, spring, and summer compared to the fall, according to the model's estimated positive coefficients, as displayed in **Table (1)**. That is, the chances of producing milk were 1.232, 1.282, and 1.195 times greater for cows that calved in the winter, spring, and summer, respectively, than for cows that calved in the fall.

Upon analysis, it was shown that the 305-milk yield had a statistically significant ($P < 0.001$) impact on milk production. Milk yield was significantly correlated negatively with 305-MY; a one-unit increase in 305-MY lowers the log chances of decreased milk production by 1.402 and 0.025 relative to the expected odds. Meanwhile, the assessment of AFC exhibited a statistically significant ($P < 0.05$) impact on milk output. Increasing AFC by one unit reduces the log chances in favor of a lower milk yield by 0.032 and 0.097 in estimated odds. AFC was shown to be a negative significant predictor of milk production. While, the results of the CI showed that it affected milk production in a statistically significant ($P < 0.001$) way. A unit increase in CI reduces the log chances in favor of a lower milk yield by 0.298 and 0.074 in stated odds, respectively. CI was found to be a significant negative predictive tool of milk yield.

As shown in **Table 1**, DIM contributed significantly (P <0.001) to the production of milk. DIM was a strong negative predictor of milk production; an increase in DIM per unit reduced the stated odds by 0.042 and the log odds in favor of a decreased milk yield by 0.875.On the other hand, the examination of DO revealed that they contributed statistically (P <0.001) to milk output. Increases in DO per unit increased the log odds in favor of decreased milk yield by 0.418 and the expected odds by 1.618, indicating that DO was a positive significant indicator of milk production.

Upon inspection, it was shown that the DP significantly (P <0.001) increased milk output. The increasing of DP per unit lowered the log chances in favor of decreased milk yield by 0.343 and 0.071 in reported odds, DP was found to be a negative significant predictor of milk production. Furthermore, the evaluation of SPC suggested that it contributed statistically (P <0.05) to milk output. Milk yield was positively and significantly predicted by SPC, with higher SPC per unit, increasing the expected odds by 0.056 and 1.058 in favor of lower milk output.

For each cow, the OLR model was built using a multiple predictors:

$$\sum_{k=1}^{K} b_{\kappa}X_{\kappa} = [-0.695\mathrm{P2} - 0.580\mathrm{P3} - 1.000\mathrm{P4} - 0.619\mathrm{P5} - 0.546\mathrm{P6} + 0.702\,\mathrm{P7} - 15.223\mathrm{P8} + 0.209\mathrm{Winter} + 0.249\mathrm{Spring} + 0.178\mathrm{Summer} - 1.402\;305 - \mathrm{MY} - 0.032\,\mathrm{AFC} - 0.298\mathrm{CI} - 0.875\mathrm{DIM} - 0.343\mathrm{DP} + 0.481\mathrm{DO} + 0.056\mathrm{SPC}\,]$$

3. Evaluation of classification technique:

The suggested OLR model's robustness is validated through classification performance measures like SEN, PREC, PPV, NPV, F1 score, AUC value along with ACC (**Tables 2)**. The ACC was determined to be 56% for the training set and 58% for the testing set. OLR's models for the training set have SEN, SPEC, PPV, NPV, and F1 values in the 0.42-0.86, 0.70-0.89, 0.43-0.77, 0.69-0.93, and 0.43-0.81 ranges, respectively. The AUC values for the training and testing sets were 0.69 and 0.68, respectively.

Table 1. *The ordered logit model for investigating the factors affecting the milk yield of Holstein-Friesians.*

	Coefficients: β (SE)	P value	OR
Threshold parameters			
Moderate milk yield threshold	5.488 (0.034)	< 0.001**	-
High milk yield threshold	5.303 (0.033)	< 0.001**	-
Location parameters			
P1(base category)			
P2	-0.695 (0.006)	< 0.001**	0.049
P3	-0.580 (0.008)	< 0.001**	0.056
P4	-1.000 (0.011)	< 0.001**	0.037
P5	-0.619 (0.016)	< 0.001**	0.054
P6	-0.546 (0.024)	0.002*	0.058
P7	0.702 (0.039)	0.007*	2.017
P8	-15.223 (0.002)	< 0.001**	0.001
Fall (base category)			
Winter	0.209 (0.006)	< 0.001**	1.232
Spring	0.249 (0.008)	< 0.001**	1.282
Summer	0.178 (0.006)	< 0.001**	1.195
305-MY	-1.402 (0.003)	< 0.001**	0.025
AFC	-0.032 (0.002)	0.02*	0.097
CI	-0.298 (0.003)	< 0.001**	0.074
DIM	-0.875 (0.009)	< 0.001**	0.042
DO	0.481 (0.009)	< 0.001**	1.618
DP	-0.343 (0.003)	< 0.001**	0.071
SPC	0.056 (0.003)	0.003*	1.058

**Coefficient is significant at a 0.001 level of significance *(P <0.001)*
*Coefficient is significant at a 0.05 level of significance *(P <0.05)*
(P1-P8) = Parity, (Spring-Summer-Fall-Winter) = Season, , 305-MY= 305-day milk yield, AFC= Age at first calving, CI= Calving interval, DIM= Days in milk, DO= Dry open, DP= Dry period, SPC= Services per conception. Baseline categories: Parity (P1), Season (Fall).

Table 2. *The model measurements of performance in both the train and test sets depend on OLR.*

Metrics	Train set			Test set		
	High	Medium	Low	High	Medium	Low
ACC	58			56		
SEN	0.86	0.42	0.51	0.84	0.38	0.51
SPEC	0.89	0.70	0.77	0.89	0.70	0.75
PPV	0.77	0.55	0.43	0.76	0.50	0.42
NPV	0.93	0.75	0.69	0.93	0.75	0.66
F1	0.81	0.43	0.53	0.80	0.40	0.50
AUC	0.69			0.68		

ACC Overall accuracy, *SEN* Sensitivity, *SPEC* Specificity, *PPV* Positive predictive value, *NPV* negative predictive value, *AUC* Area under curve.

Discussion

The current study categorized and predicted Friesian milk yield level using OLR. The POA is critical for assuring the reliability and interpretability of OLR conclusions when evaluating data with ordered categories. The Brant test demonstrated that the POA was met. The results of the current Brant test were consistent with previous studies on the POA (**Hosmer *et al.* (2013)**; **Dohoo *et al.* (2003)**; **Agga and Scott (2015)**; **Moawed and Abd El-Aziz (2022)**), allowing for proceeding with OLR analysis.

The threshold parameters were obviously statistically significant, implying that the associated cut-off point between milk yield levels was relevant and not influenced by chance. Therefore, this suggested that there was genuine change in the underlying relationship between predictors and milk yield at that exact time and confirmed the robustness of OLR to the existing datasets. This conclusion was consistent with **Moawed and Osman (2018)**, **Akkuş *et al.* (2019)**, **Akkus and Sevinc (2020)**, **Manu *et al.* (2020)**, and **Moawed and Abd El-Aziz (2022)**.

The categories of the lactation order, except 7th parity in comparison to the 1st lactation order showed a statistically significant decrease in milk supply due to the negative signs of the estimated coefficients. Conversely, the model coefficient for 7th parity was positive and substantially affected milk yield with a p-value of 0.007. This means that with every unit increase in 7th parity, the expected probability in favor of a lower degree of milk yield will increase by a multiplicative factor of 2.017. Previous studies that examined OLR for estimating factors impacting milk yield, such as the study of **Akkus and Sevinc (2020)**, who reported that the 4th parity level significantly affected milk yield. While **Moawed and Abd El-Aziz (2022)** showed that cows on the 2nd and 4th parities produced less milk than cows on the 6th lactation phase. However, it was demonstrated by **M'hamdi *et al.* (2012)** and **Akkuş *et al.* (2019)** that cows on their 1st parity produced less milk than cows on their 6th lactation order. On the other hand, these findings differed from those of **Habib *et al.* (2003)**, who found that lactation number had no effect on milk production ($P > 0.05$).

The positive coefficients predicted by the model for the winter, spring, and summer seasons suggested that milk yield increased considerably compared to the fall season. The present result coincides with the findings of **(Mundan *et al.*, 2020)**, whereas the findings of **Akkuş *et al.* (2019)**, which indicated that the most milk was produced in the fall and the least in the winter, were in conflict with our findings. Moreover, summertime milk yield at dairy farms was higher than wintertime yield **(Lavon, 2018, Abo-Gamil *et al.*, 2021)**.

The results of this study showed that the 305-day milk yield was a valuable commonly used predictor of dairy cow milk output. Further investigation revealed that these results were similar to those of **Quist *et al.* (2007)**, who demonstrated that completed 305-day milk yields were a more accurate indicator of milk supply. Nevertheless, **Abdelrahman *et al.* (2020)** found that the 305-day milk production was the most important predictor for differentiation across calving seasons.

Additionally, the age at first calving was found to be a reliable predictor of milk yield in dairy cows. The results of a comparable studies by **Eastham *et al.* (2018)**, **Akkuş *et al.* (2019)**, and **Moawed and Abd El-Aziz (2022)** that used OLR to define factors influencing milk yield in Holstein-Friesian cows provided additional support for the current findings. Nevertheless, **Yadav *et al.* (2018)** discovered that using multiple linear regressions, AFC is not a valid predictor of total milk yield in crossbred cows.

Furthermore, CI has a complicated and multidimensional association with milk yield in dairy cows. These findings contradicted those of **Remmik *et al.* (2020)**, who found that a longer initial CI leads to higher lifetime milk production.

The DO effect was one of the most valuable and financial traits that has been studied, its impact on milk production was shown to be statistically significant (P <0.001). The outcomes presented by **Temesgen *et al.* (2022)** and **Fathy *et al.* (2023b)**, as well as **Moawed and Abd El-Aziz (2022)**, were in agreement with this outcome.

The study found that DIM was a statistically significant negative contributor to milk production. This conclusion is consistent with findings of **Bajwa *et al.* (2004)** who reported the increasing trend in lactation length over the last decade, but a decrease in average lactation length in 1999 and 2000 which might be due to incomplete lactations. This is unfavorable as genetic progress has not been made in the herd. Other studies (**Osman *et al.* (2013)**, and **Akkuş *et al.* (2019)**, and **Moawed and Abd El-Aziz (2022))** indicated that lactation length had an impact on the volume of milk produced by Holstein cows.

A statistically significant negative contribution of DP to milk production was identified. This result was consistent with (**Mansfeld *et al.* (2012)**; **Kok *et al.* (2017)**; **Kok *et al.* (2021)**; **Boustan *et al.* (2021)**; and **Moawed and Abd El-Aziz (2022)**) who studied the effect of DP on milk output.

The study found that total services per conception significantly contributed to milk production, with SPC being a positive predictor of milk yield, which supported by **Abass (2010),** who found a correlation between an increase in the level of milk yield of the cow up to 9,200.61 kg and an increased number of SPC for up to six services per cow.

Conclusion

The OLR was shown to be more accurate than standard linear regression models in detecting the non-linear and ordered character of milk production data. Although OLR provides insightful information, it's important to be aware of its limits. Additional investigation into the incorporation of other cow attributes and environmental variables may enhance the precision of the prediction. Overall, this study proved the advantageous effects of using OLR to predict milk production in Friesian cows. Ultimately, such developments will help to ensure sustainable and effective milk production, which will benefit both producers and consumers.

References

Abass, H. 2010. *Genetic studies on milk production traits and breeding efficiency in Friesian cows.* MSc thesis, Benha University; Benha, Egypt.

Abdelrahman, H. A., Kamel, E. R. and Hassan, F. A. (2020). Application of linear discriminant model to evaluate the association between milk production, reproductive performance, and calving season in dairy cattle. *Benha Veterinary Medical Journal, 39*(2), 138-141.

Abdullah, F. A., Ali, J. and Noor, M. S. Z. (2021). *Factors influencing the adoption of innovation in beef cattle farming: A study in Peninsular Malaysia.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

Abo-Gamil, Z., El-Qaliouby, H., Manaa, E. and Ramadan, S. (2021). Estimation of Non-genetic Parameters Affecting Total Milk Yield and Occurrence of Mastitis in Holstein Friesian Dairy Cows. *Benha Veterinary Medical Journal, 40*(2), 184-188.

Agga, G. E. and Scott, H. M. (2015.) Use of generalized ordered logistic regression for the analysis of multidrug resistance data. *Preventive veterinary medicine, 121*(3-4), 374-379.

Akkus, O. and Ozkoc, H. (2012). A comparison of the models over the data on the interest level in politics in turkey and countries that are members of the European Union: Multinomial or ordered logit model? *Research Journal of Applied Sciences, Engineering Technology, 4*(19), 2646-2657.

Akkuş, Ö., Sevinc, V., Takma, Ç. and Güneri, Ö. I. (2019). Estimation of parametric single index ordered logit model on milk yields. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi, 25*(5), 597-602.

Akkus, Q. and Sevinc, V. 2020. Use of ordered logit model with time series data for determining the factors affecting the milk yield of Holstein Friesians. *Indian Journal of Animal Research, 54*(4), 488-493.

Bajwa, I., Khan, M., Khan, M. and Gondal, K. (2004). Environmental factors affecting milk yield and lactation length in Sahiwal cattle. *Pakistan Veterinary Journal, 24*(1), 23-27.

Boustan, A., Vahedi, V., Abdi Farab, M., Karami, H., Seyedsharifi, R., Hedayat Evrigh, N., Ghazaei, C. and Salem, A. (2021). Effects of dry period length on milk yield and content and metabolic status of high-producing dairy cows under heat stress. *Tropical Animal Health Production, 53*, 1-8.

Dohoo, I. R., Martin, W. and Stryhn, H. E. (2003). *Veterinary epidemiologic research*. Charlottetown, P.E.I.: University of Prince Edward Island.

Eastham, N. T., Coates, A., Cripps, P., Richardson, H., Smith, R. and Oikonomou, G. (2018). Associations between age at first calving and subsequent lactation performance in UK Holstein and Holstein-Friesian dairy cows. *PloS one*, *13*(6), e0197764.

El-Kasrawy, N. I., Swelum, A. A., Abdel-Latif, M. A., Alsenosy, A. E.-W. A., Beder, N. A., Alkahtani, S., Abdel-Daim, M. M. and Abd El-Aziz, A. H. (2020). Efficacy of different drenching regimens of gluconeogenic precursors during transition period on body condition score, production, reproductive performance, subclinical ketosis and economics of dairy cows. *Animals, 10*(6), 937.

Fao (2023). Food outlook.Biannual Report on Global Food Markets. Italy, Rome.

Fathy, A., Elsayed, D. H., Ibrahim, I. M. and El-Azzazi, F. E. (2023a). Multivariate Analysis on Reproductive and Productive Traits of Egyptian Buffaloes. *Journal of Advanced Veterinary Research, 13*(3), 394-399.

Fathy, A., Hassan, R. and Soliman, E. (2023b). Productive and reproductive traits and clinical epidemiological measures in discriminant analysis between baladi buffaloes and crossbred. *Adv. Anim. Vet. Sci., 11*(1), 45-56. doi:http://dx.doi.org/10.17582/journal.aavs/2023/11.1.45.56

Garamu, K. (2019). Significance of feed supplementation on milk yield and milk composition of dairy cow. *Dairy Vet Sci J, 13*(2), 555860.

Greene, W. H. (2000). *Econometric analysis*: New York University, Prince Hall, Upper Saddle River, New Jersey

Grzesiak, W., Lacroix, R., Wójcik, J. and Blaszczyk, P. (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science, 83*(2), 307-310.

Habib, M., Bhuiyan, A., Bhuiyan, M. and Khan, A. (2003). Performance of Red Chittagong Cattle in Bangladesh Agricultural University Dairy Farm. *Bangladesh Journal of Animal Science, 32*(1-2), 101-108.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.): John Wiley & Sons.

Kok, A., Van Hoeij, R., Kemp, B. and Van Knegsel, A. (2021). Evaluation of customized dry-period strategies in dairy cows. *Journal of Dairy Science, 104*(2), 1887-1899.

Kok, A., Van Knegsel, A., Van Middelaar, C., Engel, B., Hogeveen, H., Kemp, B. and De Boer, I. (2017). Effect of dry period length on milk yield over multiple lactations. *Journal of Dairy Science, 100*(1), 739-749.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational structural biotechnology journal, 13*, 8-17.

Kwak, S. K. and Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean journal of anesthesiology, 70*(4), 407-411.

Lavon, Y. (2018). The effects of summer on milk production and its components, SCC and conception rate. *Engormix.Israel Cattle Breeders Association (ICBA).*

M'hamdi, N., Bouallegue, M., Frouja, S., Ressaissi, Y., Brar, S. K. and Hamouda, M. B. (2012). Effects of environmental factors on milk yield, lactation length and dry period in Tunisian Holstein cows. *Milk Production-An Up-to-Date Overview of Animal Nutrition, Management and Health.* IntechOpen.

Macciotta, N. P. P., Vicario, D., Pulina, G. and Cappio-Borlino, A. (2002). Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *Journal of dairy science, 85*(11), 3107-3114.

Malmquist, S. and Rykatkin, O. (2023). *Identification, investigation and prediction of post-COVID phenotypes: Using Cluster analysis and Ordinal logistic regression to determine severity of post-COVID.* Uppsala University.

Mansfeld, R., Sauter-Louis, C. and Martin, R. (2012). Auswirkungen der Länge der Trockenstehzeit bei Milchkühen auf Leistung, Gesundheit, Fruchtbarkeit und Kolostrumqualität. *Tierärztliche Praxis Ausgabe G: Großtiere/Nutztiere, 40*(04), 239-250.

Manu, H., Meena, H. and Priyanka, B. (2020). Likelihood of consuming cloned animal products: ordered logistic regression model. *Intentional Journal of Current Microbiology Applied Sciences, 9*(1), 479-485.

Martinson, K., Coleman, R., Rendahl, A., Fang, Z. and Mccue, M. (2014). Estimation of body weight and development of a body weight score for adult equids using morphometric measurements. *Journal of animal science, 92*(5), 2230-2238.

Minka, T. (2001). Algorithms for maximum-likelihood logistic regression. *Statistics Tech Report, 758*.

Moawed, S. and Osman, M. (2018). Dimension reduction of phenotypic yield and fertility traits of Holstein-Friesian dairy cattle using principle component analysis. *International journal of veterinary science, 7*(2), 75-81.

Moawed, S. A. and Abd El-Aziz, A. H. (2022). The estimation and interpretation of ordered logit models for assessing the factors connected with the productivity of Holstein–Friesian dairy cows in Egypt. *Tropical Animal Health Production, 54*(6), 345.

Mundan, D., Zonturlu, A. K., Öztürk, Y., Akkuş, T. and Kaçar, C. (2020). Effect of calving season, calving year and lactation number on the milk yield traits in Holstein cows raising in Şanlıurfa. *Turkish Journal of Agriculture-Food Science Technology, 8*(2), 313-317.

Osman, M., El-Bayomi, K. and Moawed, S. (2013). Estimation of heritabilities, genetic correlations, phenotypic correlations and genetic trends for production and reproduction traits of Holstein-Friesian dairy cattle using sire model *Suez Canal Veterinary Medical Journal (SCVMJ), 18*(1), 115-128. doi:10.21608/scvmj.2013.78285

Ozawa, M., Inoue, M., Uchida, K., Chambers, J. K., Takeuch, Y. and Nakayama, H. (2019). Physical signs of canine cognitive dysfunction. *Journal of Veterinary Medical Science, 81*(12), 1829-1834.

Peng, J. 2019. Ordered logistic regression analyze the effect of elements in serum and seminal plasma on abnormal sperm rate in boars. *Journal of Animal Science, 97*(Suppl 3), 117.

Piles, M., Díez, J., Del Coz, J., Montañés, E., Quevedo, J., Ramón, J., Rafel, O., López-Béjar, M. and Tusell, L. (2013). Predicting fertility from seminal traits: Performance of several parametric and non-parametric procedures. *Livestock Science, 155*(1), 137-147.

Quist, M., Leblanc, S., Hand, K., Lazenby, D., Miglior, F. and Kelton, D. (2007). Agreement of predicted 305-day milk yields relative to actual 305-day milk weight yields. *Journal of dairy science, 90*(10), 4684-4692.

Remmik, A., Värnik, R. and Kask, K. (2020). Impact of calving interval on milk yield and longevity of primiparous Estonian Holstein cows. *Czech Journal of Animal Science, 65*(10), 365-372.

Sessa, J. and Syed, D. (2016). *Techniques to deal with missing data.* Paper presented at the 5th international conference on electronic devices, systems and applications (ICEDSA).

Shittu, A., Zaharadeen, M., Fasina, F., Umaru, M. and Ahmed, A. (2014). Classification of slaughtered animals and estimation of body condition scores during rainy season in Sokoto abattoir. *Sokoto Journal of Veterinary Sciences, 12*(2), 31-40.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing management, 45*(4), 427-437.

Suleiman, S. and Badamsi, S. (2019). Effect of Multicollinearity in Predicting Diabetes Mellitus Using Statistical Neural Network. *European Journal of Advances in Engineering Technology, 6*(6), 30-38.

Susanto, A., Hakim, L. and Nurgiartiningsih, V. (2019). *Environment (Year and Season of Birth) Effects on First-Lactation Milk Yield of Dairy Cows.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

Tabachnick, B. and Fidell, L. (2001). Using multivariate statistics *Boston: Allyn & Bacon*.

Temesgen, M. Y., Assen, A. A., Gizaw, T. T., Minalu, B. A. and Mersha, A. Y. (2022). Factors affecting calving to conception interval (days open) in dairy cows located at Dessie and Kombolcha towns, Ethiopia. *PloS one*, *17*(2), e0264029.

Topuz, D. (2021). Lactation milk yield prediction with possibilistic logistic regression analysis. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi, 27*(5).

Ursino, M. (2014). Ordinal data: a new model with applications. *PhD Thesis, Politecnico di Torino, Italy*.

Valchev, V., Marinov, I. and Angelova, T. (2020). Relationship Between Age at First Calving and Longevity and Productive Life in Holstein Cows. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 68*(5), 867–87.

Vasconcelos, J., Martins, A., Petim-Batista, M., Colaço, J., Blake, R. and Carvalheira, J. (2004). Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *Journal of dairy science, 87*(8), 2591-2598.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M. and Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one*, *12*(4), e0174944.

Yadav, A., Shukla, A. and Snehdeep, S. M. (2018). Prediction of future milk yield from early lactation traits in cross-bred cattle using multiple linear regression models. *Plann Pract Res, 13,* 613-6.

Zhang, F., Shine, P., Upton, J., Shaloo, L. and Murphy, M. D. (2018). A Review of Milk Production Forecasting Models: Past & Future Methods. *ResearchGate*.

الملخص العربي

يعتبر إنتاج اللبن من أهم مؤشرات الإنتاجية فى مزارع الألبان. ومن ثم، فإن التنبؤ الدقيق بإنتاج الحليب أمر بالغ الأهمية لتحسين إدارة مزارع الألبان وربحيتها. و يُعد الانحدار اللوجيستى الترتيبى أداة إحصائية هامة تستخدم لدراسة العلاقات بين متغير تابع ترتيبى ومجموعة من المتغيرات المستقلة . وبالتالى تهدف هذه الدراسة إلى تقييم فعالية الانحدار اللوجيستى الترتيبى فى دراسة العوامل التى تؤثر على إنتاج اللبن فى الأبقار الحلوب لتصنيف إنتاج الحليب في أبقار الفريزيان والتنبؤ به إلى فئات منخفضة (4500 كجم)، متوسطة (4500-7500 كجم)، ومرتفعه الانتاجية (< 7500 كجم). و قد تم تجميع البيانات عن إنتاج اللبن ومجموعة من العوامل المحتملة المؤثرة عليه من عدد **3739** بقرة حلوب .وجدت الدراسة أن الانحدار اللوجيستى الترتيبى كان قادرًا على تفسير جزء كبير من التباين فى إنتاج اللبن وتم تحديد مجموعة من العوامل المؤثرة على إنتاج اللبن. أشارت النتائج إلى أن عدد المواسم و عدد الأيام التي مرت منذ الولادة وانتاج اللبن فى 305 يوم كانت أهم العوامل ($p<0.05$) التي تؤثر على تصنيف البيانات.كما أظهر نموذج الانحدار اللوجيستى الترتيبى ملاءمة مُرضية في التنبؤ بفئات إنتاج الحليب، حيث أظهر دقة كبيرة (56%) ومساحة تحت المنحنى تساوي 0.69. وبالتالى أظهرت هذه الدراسة أن الانحدار اللوجيستى الترتيبى أداة فعالة لدراسة العوامل التى تؤثر على إنتاج اللبن فى الأبقار الحلوب .يمكن استخدام هذه الأداة لتحديد العوامل التى تؤثر بشكل كبير على إنتاج اللبن، مما يساعد على تحسين الإنتاجية وكفاءة الإنتاج فى مزارع الألبان.

بناءً على نتائج هذه الدراسة، يوصى بما يلى: (1) استخدام الانحدار اللوجيستى الترتيبى لدراسة العوامل المؤثرة على إنتاج اللبن فى مزارع الألبان.(2) التركيز على العوامل الوراثية والبيئية والغذائية التى ثبت أنها تؤثر بشكل كبير على إنتاج اللبن. (3) اتخاذ الإجراءات اللازمة لتحسين هذه العوامل فى مزارع الألبان.