

# Statistical Assessment of Some Factors Affecting Calving Interval by Using Ordinal Logistic Regression in Holstein Cows

Fatma D. M. Abdallah<sup>1</sup>, Eman A. Abo Elfadl<sup>2,\*</sup>

<sup>1</sup>Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt

<sup>2</sup>Department of Animal Husbandry and Development of Animal Wealth, Faculty of Veterinary Medicine, Mansoura University, Egypt

---

**Abstract Background & objectives:** Calving interval considered an important trait throughout Holstein dairy cow's life. There are many risk factors which have a great impact on it. In the line with this consideration, the purpose of this study is to apply ordinal logistic regression model to estimate the effect of these risk factors on calving interval. **Methods:** Ordinal logistic regression analysis was used to estimate the odds ratio (OR) and probability of Holstein dairy cows conception for 3400 lactation records from Dina farms company, Egypt. The data was collected over a period extended from 1998 to 2010. The logit link function was used to predict the probability of occurrence of pregnancy using SPSS version 20.0, USA. **Results:** The odds ratio showed that the likelihood of pregnancy in cows with different parities was 0.931, 0.787, 0.634 and 1.000 for lactation order 2, 3, 4 and 5 respectively. Odds ratio of pregnancy of cows calving in winter were higher than those calving in summer, it was 1.234 and 1.000 respectively. Odds of different periods of days open were 0.586, 0.771, 0.638 and 1.000 respectively. Odds of different periods of dry period were 0.378, 0.525, 0.545 and 1.000 respectively. **Conclusions:** Findings showed fitting of this model to the data, it also showed the ability of ordinal logistic regression to provide measure which facilitates understanding of the important risk factors affecting calving interval.

**Keywords** Calving interval, Pregnancy, Ordinal logistic regression, Odds ratio

---

## 1. Introduction

Logistic regression analysis is a regression model where the dependent variable (DV) is categorical; it is divided into binary, multinomial and ordinal logistic regression analysis, depending on the scale type and the number of categories of the dependent variable. It is also divided into univariate and multivariate logistic regression according to the number of risk factors [1].

Ordinal logistic regression is a new statistical method used to analyze ordered categorical response variables. It is called the proportional odds model.

These models are now available to make full use of ordered data but they are not used on a large scale [2].

The term logistic regression analysis originated from logit transformation, which is applied to the dependent variable and this causes certain differences both in estimation and interpretation [3].

Maximum likelihood method is used in logistic regression analysis to obtain coefficients, as an important step in the

analysis process before the application of logistic regression analysis [4, 5].

Logistic regression provides a measure of interest which is called odds ratio (OR). An odd is defined as the probability of an event (in this case pregnancy) occurring (p) divided by 1 - p. Odds ratios provided by multiple logistic regression (the antilogarithm of the coefficient for an independent variable) are conditional upon all other variables in the model.

## 2. Material and Methods

### 2.1. Source of Data

Data of 3400 lactation records for Holstein dairy cows were collected from records of Holstein Friesian cattle maintained at Alexandria-Copenhagen dairy farm, 76 km Alexandria-Cairo desert road during the period extended from 1998 to 2010.

### 2.2. Studied Variables

**Days open (DO):** The period per days between calving and next conception in cows, it is also called calving-to-conception interval.

---

\* Corresponding author:

emmy\_f1984@yahoo.com (Eman Abo Elfadl)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

**Calving interval (CI):** The average time interval per days between two successful calving.

**Dry period (DP):** The period during the lactation cycle when the cow is not lactating, i.e. the period between the end of the lactation and the beginning of the next one.

**Season and lactation order (parity):** Number of lactation season.

#### The data prepared and coding:

First step of this analysis was to change Continuous data to categorical data including calving interval, days open and dry period.

- ❖ Data of calving interval were categorized into three categories ( $\leq 365$  days = **1**, 366-424 days = **2** and  $< 424$  days = **3**).
- ❖ Data for effect of days open were categorized into four categories ( $< 90$  days = **1**, 90-129 days = **2**, 130-150 = **3** and  $< 150$  days = **4**).
- ❖ Data for dry period effects were categorized into four categories ( $< 45$  days = **1**, 45-60 days = **2**, 61-80 days = **3** and  $< 80$  days = **4**).
- ❖ Data for lactation order were categorized into five categories ( $P_1$  to  $P_5$ ).
- ❖ Data for season effects were categorized into two categories (winter = **1** and summer = **2**).

All these variables were entered into the equation as independent variables except calving interval which is the dependent or the outcome variable.

### 2.3. Statistical Analysis

Multivariable logistical regression procedure was used; to determine the effect of some risk factors including (lactation order, calving season, days open and dry period) on calving interval. Rate was examined, simultaneously. The multivariable logistical regression model was constructed using the PLUM procedure of [6].

Maximum likelihood value of the baseline model, Wald statistics to test significance of variables and *Exp* (B) or exponentiated logistic coefficients were assessed [7].

Cox & Snell  $R^2$ , Nagelkerke  $R^2$  and McFadden values are different ways of prediction of the explained variance in the dependent variable by the model and are interpreted as  $R^2$  in multiple regression [8]. Hosmer and Lemeshow chi-squared goodness of fit test assess whole fitness of the logistic regression model. It is more powerful than chi-square test when predictive variables are constant variables or when there is a small sample [9].

The important measure in logistic regression is the odds ratio (OR) which is the probability of an event occurring (p) divided by  $1 - p$ . It is the antilogarithm of the coefficient for an independent variable which is equal to  $\exp(\beta)$ , or sometimes written  $e^\beta$ . Some authors use the Greek symbol  $\psi$  (psi, pronounced like "sci" in science) to refer to the odds ratio, and others use OR or O.R.

For the analysis process, independent variables divided into categories. One category was considered as the

reference category for odds ratio (OR = 1). An OR of  $>1$  indicates increased likelihood while an OR  $<1$  indicates an inverse association. The outcome for each category of independent variable was compared to reference group for that variable. A cumulative logit model approach using PLUM procedure in SPSS was used. Statistical significance for all models was defined as P value  $< 0.05$ .

The specific form of the ordinal logistic regression model:

If Y is ordinal, cumulative odds or cumulative logits is used to take ordering in consideration.

Consider  $k+1$  is the ordered categories and defined by

$$P(Y \leq i) = p_1 + \dots + p_i \quad (1)$$

$$\text{Odds } (Y \leq i) = \frac{P(Y \leq i)}{1 - P(Y \leq i)} = \frac{p_1 + \dots + p_i}{p_{i+1} + \dots + p_{k+1}} \quad (2)$$

$$\text{Logit } (Y \leq i) = \ln\left(\frac{P(Y \leq i)}{1 - P(Y \leq i)}\right), \quad i = 1, \dots, k \quad (3)$$

The cumulative logistic model for ordinal response data is calculated by

$$\text{Logit } (Y \leq i) = \alpha_i + \beta_{ij}x_1 + \dots + \beta_{im}x_m \quad i = 1, \dots, k \quad (4)$$

Where, Y is the dependent variable;  $\alpha$  is the intercept parameter and  $\beta_{ij}$  to  $\beta_{im}$  are the logistic regression coefficients (parameter estimates) for the explanatory effects ( $X_1$  to  $X_m$ ) included in the statistical model.

The cumulative odds are given by

$$\text{Odds } (Y \leq i) = \exp(\alpha_i) \exp(\beta_{ij}x_1 + \dots + \beta_{im}x_m) \quad (5)$$

This means that the odds are proportional [10].

## 3. Results

The results showed fitting of overall model to the data using  $-2 \log$  likelihood = 972.487 with a p-value 0.000\*\* which is highly significant, so the covariates have a significant effect on the probability of pregnancy as shown in Table 1.

**Table 1.** Model Fitting Information Using Logit Link Function

Model	-2Log Likelihood	Chi-Square	Sig. or P value
Intercept Only	1098.745		
Final	972.487	126.259	0.000**

The Pearson chi-square statistic and the Deviance chi-square statistic goodness of fit = 320.3 and 353.1 respectively showed p-value 0.000\*\* which is highly significant as shown in Table 2.

**Table 2.** Goodness-of-Fit Using Logit Link Function

	Chi-Square	Sig. or P value
Pearson	320.331	0.000**
Deviance	353.142	0.000**

Cox & Snell, Nagelkerke and McFadden  $R^2$  are 0.052, 0.059 and 0.025 respectively. These values predict 52%, 59% and 25% of the explained variance in the dependent variable by the model as shown in Table 3.

**Table 3.** Pseudo R-Square Logit Link Function

Pseudo R-Square	
Cox and Snell	0.052
Nagelkerke	0.059
McFadden	0.025

The logistic regression analysis demonstrated that calving interval is significantly affected ( $P < 0.05$ ) by different periods of days open, by different periods of dry period, calving season and parity number except parity number 2 and 3 showed no significant effect ( $P \geq 0.05$ ) on calving interval according to Wald statistic as shown in Table 4, it also showed that ordinal logistic model could be used for calculating the probability of conception in Friesian dairy cow in the kingdom of Saudi Arabia as mentioned by [13].

The effect of different independent variables on calving interval can be described by the estimated odds ratio.

The odd ratio (OR), which is the probability of pregnancy divided by the probability of non-pregnancy. It also explains the relative amount by which the odds of pregnancy greater than 1.00 or less than 1.00 when the value of predictor increased by one unit. ORs (95% Confidence interval) will be illustrated as in Table 4.

The odds ratio showed that the likelihood of pregnancy in cows with days open less than 90 days is 0.586 (0.485-0.709), the odds of cows which have days open from 90 to 129 is 0.771 (0.634-0.937), the odds of cows which have days open

from 130 to 150 is 0.638 (0.483-0.844) and the odds of cows which have days open more than 150 days is 1.000.

Probability of pregnancy in cows when dry period is less than 45 days is 0.378 (0.297-0.482), when dry period is from 45 to 60 days is 0.525 (0.430-0.641), when dry period is from 61 to 80 is 0.545 (0.434-0.685) and when dry period is more than 80 days is 1.000.

The odds of pregnancy in winter season is higher than in summer season, it is 1.234 (1.061-1.436) and 1.000 respectively and this with agreement of [11] and [12].

The odds of pregnancy in different lactation orders is 0.931 (0.698-1.242), 0.787 (0.587-1.057), 0.634 (0.456-0.882) and 1.000 in lactation order 2, 3, 4, and 5 respectively as in Table 4. These results were in agreement of [13].

## 4. Conclusions

Regression analysis was used as statistical methods for evaluation of some risk factors that affect calving interval in dairy cattle. Result of this study showed Regression models can be used when the dependent is represented by an ordinal variable. The relationship between explanatory variables and the outcome can be represented by the proportional odds model. Ordinal regression is a powerful tool, easy to use, and produces an interpretable parameter that summarizes the effect between groups over all levels of the outcome. It also showed the ability of ordinal logistic regression to provide measure which facilitates understanding of the important risk factors affecting calving interval, Future studies are needed for further confirmation of this results.

**Table 4.** Results of Wald Statistics and Proportional Odds Model Using Calving Interval as Response with Three Ordered Categories

Covariate	Logistic Regression Coefficient	Standard Error	P-Value (Wald Statistic)	Odds Ratio (OR)	95% Confidence Interval
Days Open (1) < 90 Days	-0.534	0.097	0.000	0.586	0.485-0.709
Days Open (2) 90 - 129 Days	-0.261	0.100	0.009	0.771	0.634-0.937
Days Open (3) 130 - 150 Days	-0.449	0.142	0.002	0.638	0.483-0.844
Days Open (4) > 150 Days	0.000			1.000	
Dry period (1) < 45 Days	-0.973	0.124	0.000	0.378	0.297-0.482
Dry Period (2) 45 - 60 Days	-0.644	0.102	0.000	0.525	0.430-0.641
Dry Period (3) 61- 80 Days	-0.607	0.117	0.000	0.545	0.434-0.685
Dry Period (4) > 80 Days	0.000			1.000	
Season (1) Winter	0.211	0.077	0.006	1.234	1.061-1.436
Season (2) Summer	0.000			1.000	
Lactation Order (1)	0.000				
Lactation Order (2)	-0.072	0.147	0.626	0.931	0.698-1.242
Lactation Order (3)	-0.239	0.150	0.112	0.787	0.587-1.057
Lactation Order (4)	-0.456	0.168	0.007	0.634	0.456-0.822
Lactation Order (5)	0.000			1.000	

---

## REFERENCES

- [1] Stephenson B., "Binary response and logistic regression analysis". [www.public.iastate.edu/~stat415/stephenson/stat415\\_chapter3.pdf](http://www.public.iastate.edu/~stat415/stephenson/stat415_chapter3.pdf). adresinden 22 Kasım 2008 tarihinde edinilmiştir, 2008.
- [2] Bender, R., and Grouven, U., "Ordinal logistic regression in medical research". *Journal of the Royal College of Physicians of London*, 31:5, 1997.
- [3] Hair J. F., Black W. C., Babin B., Anderson R. E., and Tatham R. L., "Multivariate data analysis", 6th ed., Upper Saddle River, NJ: Prentice-Hall, 2006.
- [4] Menard S., "Applied logistic regression analysis". Thousand Oaks, CA: Sage, 1995.
- [5] Myers R., "Classical and modern regression with applications", 2nd ed., Boston, MA: Duxbury, 1990.
- [6] SPSS, "Statistical Package for Social Sciences". Release 20.0 versions. SPSS Inc. USA, 2006.
- [7] Agresti A., "An introduction to categorical data analysis", New York: John Wiley and Sons, 1996.
- [8] Field A., "Discovering statistics using SPSS", 2nd ed., London: Sage, 2005.
- [9] Hosmerand D. W. and Lemeshow S., "Applied logistic regression", John Wiley & sons, INC. NY., 2002.
- [10] McCullough P., "Regression models for ordinal data (with discussion)", *J. R. Stat. Soc. B.*, 42:109-42, 1980.
- [11] Badinga, L., Collier, R. J., Thatcher, W. W., and Wilcox, C. J., "Effect of climatic and management factors on conception rate of dairy cattle in subtropical environment", *J. Dairy. Sci.*, 68:78-85, 1985.
- [12] Ali, A. K. A., A L-Haidary, A., Al-Shaikh, M.A., Gamil, M. H., and Hayes, E., "Effect of day open on lactation curve of Holstein cows in Saudi Arabia". *Asian-Aus. J. Anim. Sci.*, 13:227-286, 1999.
- [13] Ali, A. K. A., AL Essa, A. A., Alshaikh, M. A., Aljumaah, R. S., Al-Haidary, A. A. and Alkraidees, M. S., "Odds ratio and probability of conception of Holstein Friesian dairy cows in the kingdom of Saudi Arabia". *Asian-Aust. J. Anim. Sci.*, Vol.18, No. 3: 308-313, 2005.