

Parametric Survival Models for Predicting of Pregnancy in Friesian Cattle

Fatma D. M. Abdallah¹, Eman A. Abo Elfadl^{2,*}

¹Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Egypt

²Department of Animal Husbandry and Development of Animal Wealth, Faculty of Veterinary Medicine, Mansoura University, Egypt

Abstract Background & objectives: This work was performed to apply different parametric survival models instead of nonparametric ones for predicting pregnancy in Friesian dairy cattle by using days open as a time variable. The models used were exponential, normal, log-normal, Weibull, logistic, log logistic and smallest extreme value. Data for present study were obtained from animal records belongs to different Dakhalia farms (n = 1842) Covering the period between 2009 and 2011. Variables included in this study were days open as a dependent variable and the independent variables were: age at calving, dry period, calving interval, season, and lactation order. The survival time data were modeled by using life data regression procedure of Statgraphics statistical package program. Model parameters were estimated and comparisons among models were done based on Akaike Information Criterion (AIC). The Weibull model was the best option for evaluating time till occurrence of pregnancy where it had the smallest (AIC) value. It also showed a good fit with the studied data. This Weibull survival model can be used to predict the length of time from calving to conception in Friesian. But the Logistic was not appropriate to describe the dataset.

Keywords Days open, Survival analysis, Parametric models, Censored data and Akaike Information Criterion (AIC)

1. Introduction

There are many factors influencing the fertility in dairy cattle. Days open is one of these factors which play an important role in reproductive and productive efficiency of individual cows (Arthur *et al.*, 2001). Days open is the number of days between the most recent calving and conception (calving conception interval). Standard statistical techniques such as multivariable regression analyses used to describe and quantify the effect of factors influencing days open in dairy cattle. Survival analysis is an important area in statistics as it gives a number of advantages over other regression techniques that it can solve the problem of censored observations (Kleinbaum and Klein, 2005). Censoring means that the animal does not achieve the expected event during the study period. Datasets in survival analysis are usually non-negative, continuous and its specific survival function. A survival function can be detected by adjusting parametric models, which is usually nonlinear models, instead of the usual non-parametric (Kaplan-Meier) and Cox proportional hazard model which is a semi-parametric (Colosimo and

Giolo, 2006).

The use of censored data provides less biased estimates for factors affecting reproductive performance (Del *et al.*, 2005; Del *et al.*, 2006) and for this reason, survival analysis is the recommended method for analysis of dairy cow reproductive data.

Since survival data are usually asymmetric, the assumption of normality of the data is not essential. Therefore, exponential, normal, log-normal, Weibull, logistic and log logistic distributions are generally used for these data (Collet and Kimber, 2011).

2. Material and Methods

Data for present study were obtained from animal records belongs to different Dakhalia farms (n = 1842) Covering the period between 2009 and 2011. The data set were fitted to different survival parametric models. The animals which became pregnant at the study period are considered complete data.

The following categories of records were considered as censored data:

- Cows still non pregnant at the end of the study period;
- Cows sold to another herd.
- Cows culled from the herd throughout the breeding period for reasons apart from reproductive failure.
- Cows with days open, or interval between last calving and the end of data, longer the study period were

* Corresponding author:

emmy_f1984@yahoo.com (Eman A. Abo Elfadl)

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

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

considered censored (Roxstrom *et al.*, 2003).

Variables under study were days open is a dependent variable. Season of calving (categorical independent variable) is coded as (1) in case of summer season and (summer season is omitted) and (2) in case of winter season. Calving interval, previous dry period and age at calving are continuous independent variables. Lactation order (categorical independent variable) is coded as: lactation order 1 = (1), lactation orders 2 = (2), lactation order 3 = (3), lactation order 4 = (4) and lactation order 5 and over = (5). (Lactation order 5 and over is omitted as a reference group).

2.1. Statistical Data Analysis

Multivariate regression or parametric survival analysis was performed using several survival modeling methods. Several accelerated failure time parametric models are exponential, normal, log-normal, Weibull, logistic, log logistic and smallest extreme value distributions. All

survival models were constructed using life data regression procedure of Statgraphics statistical package program (Statgraphics, 2015). Estimates for the parameters from different models fits estimated by using maximum likelihood and comparisons among different models were done by using Akaike information criterion (AIC) (Akaike, 1974).

Lower AIC values indicate a better model fit. The formula for AIC is:

$$AIC = -2 \log L + 2k$$

where $\log L$ is the log likelihood of the proposed model, and k is the number of model parameters.

As it is known that the core of survival analysis is survival and hazard functions. These functions are used in the process of prediction of the event under study.

Survival function, density function and hazard function for different parametric models were shown in Table (1) as they are the core of survival analysis.

Table (1). Survival function, density function and hazard function for different parametric models

Life data regression models	Survival Function $s(t)$	Hazard function $h(t)$	Density function $f(t)$
Exponential	$e^{-\lambda t}$	λ	$\lambda e^{-\lambda t}$
Logistic	$\frac{1}{1 + \exp(\frac{t-\mu}{\sigma})}$	$\frac{e^{-\frac{t-\mu}{\sigma}}}{\sigma(1 + e^{-\frac{t-\mu}{\sigma}})}$	$\frac{e^{-\frac{t-\mu}{\sigma}}}{\sigma(1 + e^{-\frac{t-\mu}{\sigma}})^2}$
Log logistic	$\frac{1}{1 + \alpha t^\gamma}$	$\frac{\alpha \gamma t^{\gamma-1}}{1 + \alpha t^\gamma}$	$\frac{\alpha \gamma t^{\gamma-1}}{(1 + \alpha t^\gamma)^2}$
Normal	$\Pr(T > t)$	$\frac{f(t)}{s(t)}$	$\Pr(T = t)$
Lognormal	$1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$	$\frac{\left(\frac{1}{t\sigma}\right)\phi\left(\frac{\log t}{\sigma}\right)}{\Phi\left(\frac{-\log t}{\sigma}\right)}$	$\frac{\exp[-(\log t - \mu)^2 / 2\sigma^2]}{2\sqrt{\pi}\sigma}$
Weibull	$\exp(-\lambda t^\alpha)$	$\lambda \alpha t^{\alpha-1}$	$\lambda \alpha t^{\alpha-1} e^{-(\lambda t)^\alpha}$
Smallest extreme value	e^{-e^t}	e^t	$\frac{1}{\beta} e^{-\frac{x-\mu}{\beta}} e^{-e^{-\frac{x-\mu}{\beta}}}$

In the above models, $s(t)$ is the survival over time (per days); λ is a scale parameter associated with the average lifetime in the exponential model. γ is a shape parameter associated with the hazard ratio; and μ and σ are scale and location parameters in case of hazard and density function. β also is the 3% scale parameter in case of smallest extreme value distribution (Colosimo and Giolo, 2006; Lee and Wang, 2003).

2.1.1. Parametric Survival Model

The parametric regression model can also be used to test the relationship between the time variable (days open) and the explanatory variables. Generally any distribution defined for $t \in [0, \infty]$ can be treated as a survival distribution. Through a logarithm transform we get $\log(t) \in (-\infty, +\infty)$ which is more suitable for modeling. For

example, a Weibull distribution which is most commonly used in survival analysis. The general regression model is given below (Rodriguez, 2005).

$$\ln t = \alpha + \sigma w$$

Where $\alpha = -\log(h(t))$ has a specific distribution.

For Weibull distribution, $\ln t = u + \beta_1 x_1 + \dots + \beta_q x_q + \sigma w$

Where,

X_1, \dots , are explanatory variables

B_1, \dots , are regression coefficients

σ is a scale parameter

W has the extreme value distribution

U is an intercept

The regression parameters will be used to interpret the direction and strength of the relationship of each explanatory on the effect of the survival time. Positive value of B_p is indicative of increased survival.

3. Results

A total of 1842 animals divided into 854 uncensored values with percent 46.36% (pregnant animals) and 988 censored values with percent 53.64% (non-pregnant).

After constructing different types of parametric survival models Weibull, normal, exponential, logistic, log-logistic, lognormal and smallest extreme value and compared their AIC (Table 2) to determine the best fit model. The Weibull model had the lowest AIC which is (10956.42) in days open dataset, indicating a better overall fit than the other models with parameters 5.2907(0.0901337). The Logistic was not appropriate to describe the dataset.

The parameters from different model fits are displayed in Table (2). Positive value of β_P is indicative of increased survival (increasing the event occurrence). The Akaike Informational Criterion (AIC) is also provided at this table.

Table (2). The parameter estimates and Akaike Informational Criterion (AIC) of different models

Distribution	Parameter (SE)	AIC
Exponential	5.5387 (0.174898)	11519.52
Normal	173.268 (16.6974)	11191.26
Lognormal	5.03434 (0.0967191)	10986.56
Weibull	5.2907 (0.0901337)	10956.42
Log-logistic	4.97963 (0.0961768)	10986.9
Smallest Extreme Value	184.239 (20.8397)	13327.32
Logistic	-	-

Generally, positive values for a beta coefficient of different variables indicate a better survival and this means that the cow has a good chance to be pregnant, and negative values give worse survival (pregnancy chance is lower) according to Gamel and McLean (1994).

Because the Weibull model had the lowest AIC score and demonstrated a good fit of the data, so this parametric model was selected as the best model.

The equation of the Weibull fitted model is

Days open = exp (5.2907 + 0.000297543 * Age at calving - 0.000875385 * Dry period + 0.000517566 * Calving interval + 0.0348342 * Season (1) + 4.30334 * Lactation order (1) + 0.0305605 * Lactation order (2) + 0.0764563 * Lactation order (3) + 0.0853323 * Lactation order (4).

As it is known that positive values for a beta coefficient depict better survival, and negative values depict worse survival. For example, the negative beta coefficient (-0.000875385) for a cow indicates that these cows are observed to have a worse chance of pregnancy compared to others with positive beta coefficients.

Variables such as age at calving, calving interval and season are associated with markedly improved survival as their beta's values are (+ 0.000297543, + 0.000517566 and + 0.0348342 respectively). Increasing in lactation orders variable may seem to imply a better chance of pregnancy for these cows due to positive values for beta coefficients as

in Table (3).

Table (3). The beta coefficients for the Weibull survival model

Covariate	Beta Coefficients
Age at calving	0.000297543
Dry period	- 0.000875385
Calving interval	0.000517566
Season (1)	0.0348342
Lactation order (1)	4.30334
Lactation order (2)	0.0305605
Lactation order (3)	0.0764563
Lactation order (4)	0.0853323
Sigma	0.514634

This Weibull equation can be used to predict pregnancy in cows depending model parameters and beta coefficients of the explanatory variables. The time of pregnancy (years or months) for each cow can be detected in each lactation order and season.

4. Conclusions

The parametric survival models can be used to model dairy data as it deals with censored values. The AIC is used to evaluate model performance, as this is one of the most widely used metrics. The Weibull model demonstrated the best performance compared to other parametric models so its equation is used in predicting process of the event under study. The Logistic was not appropriate to describe the dataset. Finally, this model is good for prediction process of all events belonged to dairy farms depending on the variables affecting these events.

REFERENCES

- [1] Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 19:716-723.
- [2] Arthur G, Noakes D, Pearson H, Parkinson T 2001. Veterinary reproduction and obstetrics. 8th Ed. Saunders, W.B. et al. (Eds.). London: England; p. 520.
- [3] Collet D, Kimber A. 2011. Modelling survival data in medical research. 3rd ed. London: Chapman & Hall; p.448.
- [4] Colosimo EA, Giolo SR.2006. Análise de sobrevivência aplicada. São Paulo: Edgar Blücher; p.392.
- [5] Del, M., Schneider, P., Strandberg, E., Ducrocq, V., Roth, A. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. J. Dairy Sci. 88: 2253-2259.
- [6] Del, M., Schneider, P., Strandberg, E., Ducrocq, V., Roth, A. 2006. Genetic evaluation of the interval from 1st to last insemination with survival and linear models. J. Dairy Sci. 89: 4903-4906.

- [7] Gamel, J.W., McLean, I.W. 1994. A stable, multivariate extension of the log-normal survival model. *Comput. Biomed. Res.* 27:148-55.
- [8] Kleinbaum DG, Klein M. 2005. *Survival analysis: a self-learning text*. 2nd ed. New York: Springer; p. 606.
- [9] Lee ET, Wang JW. 2003. *Statistical methods for survival data analysis*. 3rd ed. New York: John Wiley and Sons; p.534.
- [10] Rodriguez, G. 2005. *Parametric Survival Models*, [http://data.princeton.edu/pop509a/Parametric Survival.pdf](http://data.princeton.edu/pop509a/Parametric%20Survival.pdf).
- [11] Roxstrom, A., Ducrocq, V., Strandberg, E. 2003. Survival analysis of longevity in dairy cattle on a lactation basis. *INRA, EDP Sciences, Genet. Sel. Evol.* 35: 305-318.
- [12] Statgraphics, 2015. *Statgraphics Centurion XVII*, <http://www.statgraphics.com>.