Statistical Analysis Of Biological Survival Data

El-Bayomi Kh M [*], Rady El A [**], El-Tarabany M S [*] and Fatma D Mohammed [*]

[*] Department of Animal Wealth Development, Faculty of Vet. Medicine, Zagazig University.
[**] Department of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University.

ABSTRACT

The objective of this study is to throw the light on applications of survival analysis in veterinary and biological sciences. In veterinary sciences especially in dairy farms, there are important factors in dairy herds which have a great role in their effects on another important factor which is called days open. There are many statistical methods used to model and analyse the data under these circumstances, but here some methods of survival analysis will be used because there are two types of data in the study (complete data and incomplete or censored data). The common statistical methods cannot be used to analyse the censored data. Some methods such as Kaplan- Meier method, Log-rank method and the Cox's proportional hazard model method will be used in this study. The data were obtained from different lactation records, covering the period between 2004 and 2007. These milk records are of U.S. Holstein cows belonging to Dina farms. The result showed that based on the K-M survivorship percentiles, overall median days open of dairy cattle was at (134 days). There were a non-significant difference between seasons, and a highly significant difference between years and lactation order. The result from Cox's proportional hazard model showed that lactation order increased the chance of pregnancy. Age at calving, days in milk, and season and year of calving decreased it. The result of testing Cox's model assumptions using a Schoenfeld residuals showed no correlation between partial residuals of the variables under study and rank of days open, so the proportional hazard assumption is satisfied.

Keywords: Survival analysis, Cox proportional hazard model, Dairy cattle and Days open.

INTRODUCTION

Survival analysis is widely applied in many fields such as biology, medicine, public health, and epidemiology. A typical analysis of survival data involves the modeling of time-to-event data, such as the time until death. The time to the event of interest is called either survival time or failure time *(1)*.

There is no doubt that survival analysis will take a more prominent place among animal breeders, not only cattle breeding but also in other species *(2)*.

Survival analysis is an alternative method for analyzing reproductive traits recorded as time interval. Survival-analysis methods often are used to analyze data from dairy herds where the outcome of interest is the interval from calving to conception *(3)*.

The goal of survival analysis is to analyze positive measures describing the "width" of the interval between an origin point and an end point. The end point, is called failure, corresponded to death or culling of the animal. But the end point may also correspond to the occurrence of any type of event, e.g, recovery from a disease *(4)*.

The common statistical techniques employed to analyze survival data in public health research. Due to the presence of censoring, the data are not amenable to the usual method of analysis. The improvement in statistical computing and wide accessibility of personal computers led to the rapid development and popularity of nonparametric over parametric procedures. Nonparametric techniques include the Kaplan-Meier method for estimating the survival function and the Cox proportional hazards model to identify risk factors and to obtain adjusted risk ratios *(5)*.

The main objective of this study was to estimate the survival function using Kaplan-Meier method, to compare between survival curves using Log rank method and to evaluate the effect of season and year of calving, calving interval, days in milk, total milk yield, lactation order, dry period and age at calving on days open (calving conception interval) in dairy cattle data using Cox proportional hazard model.

MATERIAL AND METHODS

Data were collected from different lactation records, covering the period between 2004 and 2007. The data were taken from the milk records of U.S. Holstein cows belonging to Dina farms (The Modern Agriculture Development), located about 80 km in Cairo-Alexandria desert road.

Variables under study

The outcome variable (dependent) in this study was survival time variable of days open (calving conception interval).

The independent variables were season and year of calving, calving interval, days in milk, total milk yield, lactation order, dry period and age at calving.

Season and year of calving and lactation order are categorical variables.

Calving interval, days in milk, total milk yield, dry period and age at calving are continuous variables.

Statistical analysis

1. Firstly the data were tested for normality to determine the statistical method of analysis.
2. Then distribution fitting of data is also tested:

The aim of distribution fitting is to predict the probability or to forecast the frequency of occurrence of the magnitude of the phenomenon in a certain interval.

3. Non parametric and semi-parametric methods of survival analysis.

Kaplan-Meier method

The Kaplan–Meier method is the most widely used method in survival data analysis *(6)*.

The Kaplan-Meier method can be used to estimate the survival curve from the observed survival times without the assumption of an underlying probability distribution. The method is based on the basic idea that the probability of surviving k or more periods from entering the study is a product of the k observed survival rates for each period (i.e. the cumulative proportion surviving), given by the following:

$$S(k) = p_1 \times p_2 \times p_3 \times \dots \times p_k.$$

Where, p_1 is the proportion surviving the first period, p_2 is the proportion surviving beyond the second period conditional on having survived up to the second period, and so on. The proportion surviving period i having survived up to period i is given by:

$$p_i = r_i - d_i / r_i$$

Where r_i is the number alive at the beginning of the period and d_i the number of deaths within the period.

Log rank test

In statistics, the logrank test is a hypothesis test to compare the survival distributions of two samples. It is a nonparametric test and appropriate to use when the data are right skewed and censored.

Comparison of two survival curves can be done using a statistical hypothesis test called the log rank test. It is used to test the null hypothesis that there is no difference between the population survival curves (i.e. the probability of an event occurring at any time point is the same for each population). The test statistic is calculated as follows:

$$\chi^2(\log rank) = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Where the O_1 and O_2 are the total numbers of observed events in groups 1 and 2, respectively, and E_1 and E_2 the total numbers of expected events. The total expected number of events for a group is the sum of the expected number of events at the time of each event. The expected number of events at the time of an event can be calculated as the risk for death at that time multiplied by the number alive in the group. Under the null hypothesis, the risk of death (number of deaths/number alive) can be calculated from the combined data for both groups.

The total expected number of events for group 2 is calculated as:

$$E_2 = \sum_{i=1}^{k} \frac{d_i}{r_i} r_{2i}$$

Where r_{2i} is the number alive from group 2 at the time of event i. E_1 can be calculated as n – E_2, where n is the total number of events. The test statistic is compared with a χ^2 distribution with 1 degree of freedom *(7)*.

Cox's proportional hazard model

Cox regression is a well-known approach for modeling censored survival data. However, the model has an assumption of proportional hazards which requires an attention *(8)*.

The probability of the endpoint (death, or any other event of interest, e.g. recurrence of disease) is called the hazard. The event of interest here is pregnancy. The hazard is modeled as:

$$H(t)=H_0(t)\times\exp(b_1x_1+b_2x_2+b_3x_3+\ldots..+b_kx_k).$$

$$h(t,x) = h_0(t)\exp\{\beta_1X_1+\ldots.+\beta_kX_k\}.$$

Where h(t; x) is the hazard function at time t for a subject with covariate values $x_1, \ldots x_k$.

$h_0(t)$ is the baseline hazard function, i.e., the hazard function when all covariates equal zero.

exp is the exponential function ($\exp(x)= e^x$).

x_i is the i^{th} covariate (explanatory/predictor variables) in the model.

β_i is the regression coefficient for the i^{th} covariate, x_i.

Assumptions of the Cox's proportional hazard formula

Cox regression is sometimes described as semi-parametric because, although it is based on a parametric regression model, it does not make specific assumptions about the probability distribution of event times.

While no assumptions are made about the shape of the underlying hazard function, the model equations shown above do imply two assumptions. First, they specify a multiplicative relationship between the underlying hazard function and the log-linear function of the covariates. This assumption is also called the proportionality assumption. In practical terms, it is assumed that, given two observations with different values for the independent variables, the ratio of the hazard functions for those two observations does not depend on time. The second assumption of course, is that there is a log-linear relationship between the independent variables and the underlying hazard function.

Definition of the hazard ratio

Hazard Ratio is defined as

$$HR = \frac{h(t,x^*)}{h(t,x)}.$$

Where $X^* = (x_1^*, x_2^*, \ldots\ldots, x_p^*)$.

$X = (x_1, x_2, \ldots\ldots, x_p)$.

Statistical programs

Data were collected, arranged, summarized and then analyzed using different computer program **SPSS/ PC+** *(9)*.

RESULTS AND DISCUSSION

After testing the data for normality, the results indicated that the dependent variable under study was not normally distributed. The P value was (0.00**) that was a highly significant and the null hypothesis (H_0: the data follow the normal distribution) was rejected.

Distribution fitting results

There are different types of probability plots helped in determining whether days open (survival time variable) came from a particular type of distribution. After examining these plots, it is found that the data did not follow a known theoretical distribution.

The results of Kaplan-Meier and Logrank methods

The K-M curve for overall survival time and event from this study results was shown in Figure (1) (survival function plot).

The mean days open was 414.81 days. Based on K-M survivorship percentiles, in overall median days open of dairy cattle was at 134 days. These results were in agreement with *(10)* who mentioned that, the mean is generally used to describe the central tendency of a distribution, but in survival distributions the median is often better because a small number of individuals with exceptionally long or short lifetimes will cause the mean survival time to be disproportionately large or small.

Based on the K-M survivorship percentiles, median days open for the cow in the first season (summer) was 136 days and 134 days for cows in the second season (winter) Figure (2), and the means were 410.97 days and 371.65 days respectively.

There was no significant difference between seasons (summer and winter) since two samples survival test by Cox-Mantel test (Log-rank value = 0.172) resulted P- value = 0.68, this result is in agreement with the findings of *(11)*.

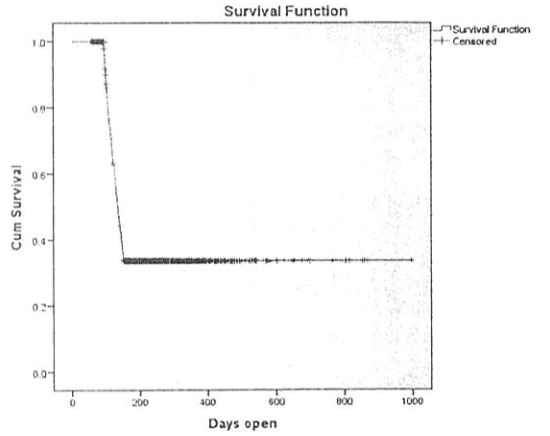
Median days open for cows in the year 2004 was 131 days, in the year 2005 was 141 days, 143 days in the year 2006, and 118 days in the year 2007 Figure (3).

The means were 505.56, 412.17, 283.48 and 147.09 days for years respectively. There was a highly significant difference between years since P-value = 0.00** and the test value was 283.93.

Median days open for cows in the first lactation order was 137 days, in the 2nd order of lactation was 127 days, in the 3rd lactation order was 130 days, in the 4th lactation order was 140 days and the fifth lactation order and over the median was 150 days as shown in Figure (4).

The means were 357.07, 313.843, 299.79, 432.47 and 440.70 days for the same lactation orders respectively. There was a highly significant difference between lactation orders since P-value = 0.00** and the test value was 86.31.

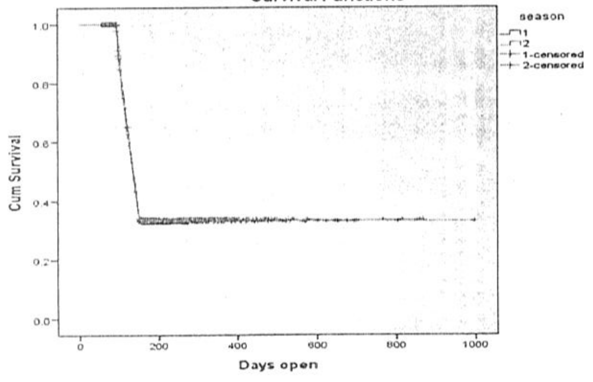
Fig. 1. Number of censored observations and the Kaplan-Meier curve for survival time (day) and event for overall model



This plot shows the time to the event on the horizontal axis, and the probability of survival on the vertical axis. The curve crosses 50% survival at the time of 134 days (median survival time). The Kaplan-Meier curve is a step function: the curve stays at a certain level until the next event occurs. At each event time, the curve drops.

Number of censored observations is 1400 value with percent 46.7%. Number of pregnant animals is 1600 value with percent 53.3%.

Fig. 2. Number of censored observations and the Kaplan-Meier curve for survival time (day) and event for season



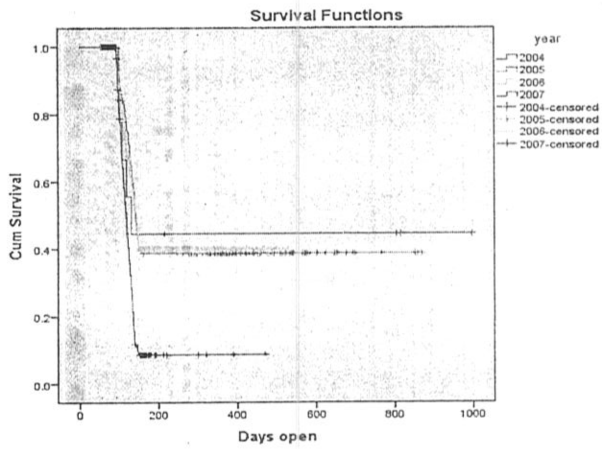
The horizontal axis shows the time to event. In this plot, drops in the survival curve occur whenever the animals become pregnant. The vertical axis shows the probability of survival.

This plot divided into 2 categories, season 1 and season 2. The curve crosses 50% survival at the time of 136 days and 134 (median survival time) respectively.

The first season: Number of censored observations is 495 values with percent 43.8%. Number of pregnant animals is 634 values with percent 56.16 %.

The second season: Number of censored observations is 896 values with percent 48.24%. Number of pregnant animals is 961 values with percent 51.75%.

Fig. 3. Number of censored observations and the Kaplan-Meier curve for survival time (day) and event for year



This plot shows the estimated survival function of days open which represent X axis and Y which represented by cumulative survival function which ranged from 0 to 1.

This plot divided into 4 categories, years of 2004, 2005, 2006, and 2007. The curve crosses 50% survival at the time of 131 days, 141, 143 and 118 (median survival time) respectively.

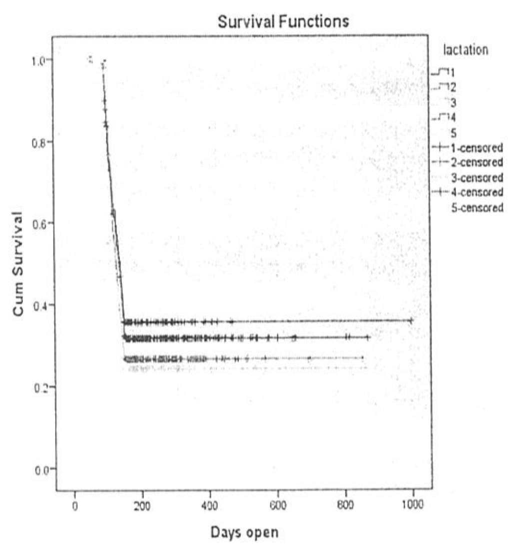
Number of censored observations at the first year is 4 values with percent 44.45%. Number of pregnant animals at the first year is 5 values with percent 55.55%.

The second year: Number of censored observations is 88 values with percent 39.47%. Number of pregnant animals is 135 values with percent 60.53%.

The third year: Number of censored observations is 1012 values with percent 49.91%. Number of pregnant animals is 1016 values with percent 50.09%.

The fourth year: Number of censored observations is 291 values with percent 39.86%. Number of pregnant animals is 439 values with percent 60.14%.

Fig.4. Number of censored observations and the Kaplan-Meier curve for survival time (day) and event for lactation order



This plot divided into 5 categories, lactation order 1, 2, 3, 4, and 5. The curve crosses 50% survival at the time of 137 days, 127, 130, 140 and 150 (median survival time) respectively.

Number of censored observations at the first lactation order is 383 values with percent 47.87%. Number of pregnant animals at the first lactation order is 417 values with percent 52.13%.

Lactation order 2: Number of censored observations is 303 values with percent 40.73%. Number of pregnant animals is 441 values with percent 59.27%.

Lactation order 3: Number of censored observations is 189 values with percent 35.86%. Number of pregnant animals is 338 values with percent 64.14%.

Lactation order 4: Number of censored observations is 105 values with percent 47.95%. Number of pregnant animals is 114 values with percent 52.05%.

Lactation order 5: Number of censored observations is 420 values with percent 59.22%. Number of pregnant animals is 290 values with percent 40.48%.

Cox's proportional hazard model results

The result of log likelihood chi-square of overall model showed P value = 0.00** meaning that there was a highly significant difference, so the variables (season and year of calving, lactation order (categorical variables), calving interval, effect of lactation length or days in milk (DIM), total milk yield, dry period, and age at calving (continuous variables) have a significant effect on the probability of pregnancy (hazard rate).

The results of fitting a failure-time regression model to describe the relationship between days open and the independent variable(s) is shown in Table (1).

Table 1. Cox's proportional hazard model (main effect model) in the studied dairy

Variable	β	SE	Wald $(\beta/SE)^2$	P- Value	Exp(β) (hazard ratio)	95 % CI for Exp(β)	
						Lower	Upper
Age at calving	-0.006526	0.002009	10.551132	0.001161**	0.993495	0.989590	0.997415
Days in milk	-0.002533	0.000389	42.489575	0.000000**	0.997470	0.996711	0.998230
Total milk yield	0.000002	0.000013	0.022689	0.880268	0.999998	0.999972	1.000024
Dry period	-0.000169	0.001048	0.025914	0.872111	0.999831	0.997780	1.001887
Calving interval	-0.000458	0.000298	2.365941	0.124009	0.999542	0.998959	1.000126
Season	-0.344793	0.063998	29.025546	0.000000**	0.708367	0.624860	0.803034
Year (1)	-0.835895	0.524376	2.541076	0.110919	0.433486	0.155105	1.211505
Year (2)	-1.568390	0.539101	8.463846	0.003623**	0.208380	0.072439	0.599433
Year (3)	-0.762216	0.555536	1. 882488	0.170052	0.466631	0.157073	1.386266
Lactation order (2)	0.491642	0.159827	9.462343	0.002097**	1.634999	1.195288	2.236466
Lactation order (3)	0.491760	0.160394	9.399996	0.002170**	1.635191	1.194100	2.239218
Lactation order (4)	0.466838	0.188425	6.138388	0.013228**	1.594943	1.102446	2.307453
Lactation order (5)	0.043037	0.171403	0.063046	0.801745ns	1.043977	0.746092	1.460795

h(t|x) = h(t|0) * exp (- 0.006526 * Age at calving - 0.002533 * Days in milk + 0.000002 * Total milk yield - 0.000169 * Dry period - 0.000458 * Calving interval - 0.344793 * (Season = 2) - 0.835895 * (Year = 2005) - 1.568390 * (Year = 2006) - 0.762216 * (Year=2007) + 0.491642 * (Lactation Order = 2) + 0.491760 * (Lactation Order = 3) + 0.466838 * (Lactation Order = 4) + 0.043037 * (Lactation Order = 5)).

This table shows the significance levels separately for the predictor variables. The test statistic used here (which is assumed to follow a normal distribution) is the Wald statistic, is just $(\beta/ SE)^2$, the square of the parameter estimate is divided by its standard deviation.

The value for Exp(β) corresponds to the relative risk. The relative risk estimates may assume only positive values.

If Exp(β), the hazard ratio > 1, the hazard rate will increase (so the expected time to the event, pregnancy, will decrease) for increasing values of the covariates. The fact that the hazard ratio for any covariate is close to one reflects its non- significance.

Age at calving, days in milk, season, year 2 (2006), lactation order (2), and lactation order (3), and lactation order (4) all are highly significant.

Total milk yield, dry period, calving interval, year 1 (2005), year 3(2007), and lactation order (5), all are not significant.

The hazard ratio for age at calving, days in milk, season at calving and year at calving is less than 1 so the chance of pregnancy reduced.

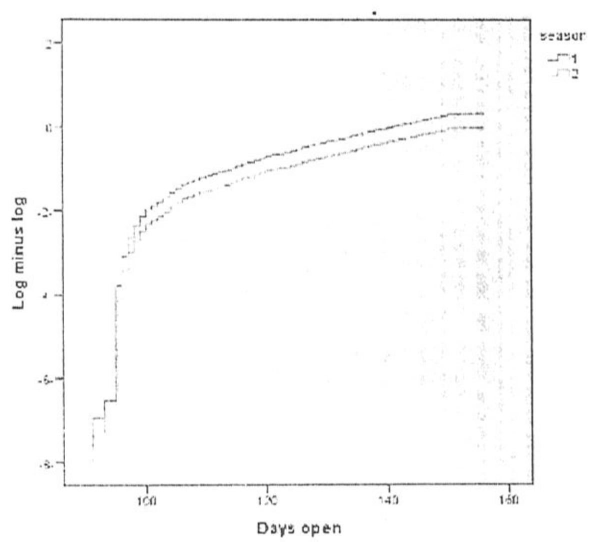
The hazard ratio for lactation order is more than 1 so the chance of pregnancy increased.

The hazard ratio for total milk yield, dry period and calving interval equal 1 so the chance of pregnancy is not affected.

Results of checking assumptions in the Cox proportional hazard regression model using Log minus log graph:

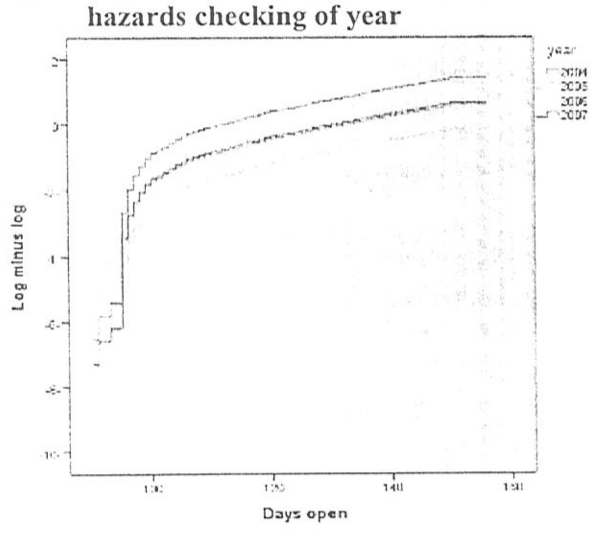
Log minus log graph is a graph used to test the assumption of the Cox model depending on its shape. It is a useful visualization of the effect of categorical variables on the survival function. This plot displays the log-minus-log of the survival function, ln(−ln(survival)), versus the survival time. This particular plot displays a separate curve or line for each category of the categorical variables. The lines should not cross each other to assess the assumption of the model as in the following Figures (5), (6) and (7).

Fig. 5. Log minus log plot for proportional hazards checking of season



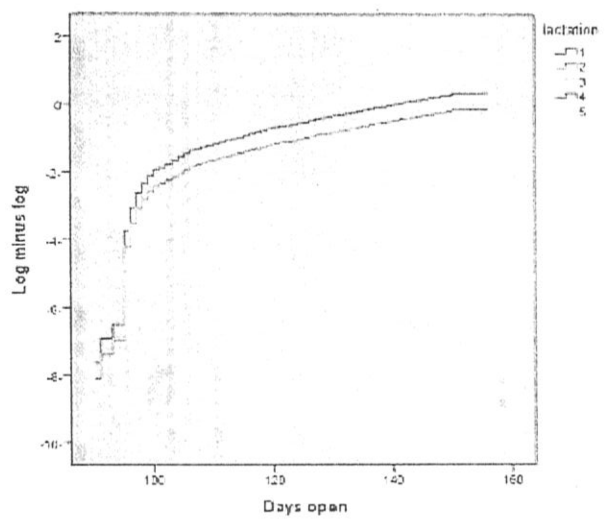
This plot displays a separate curve for each season category (approximately parallel curves) support proportional hazard assumption for season.

Fig.6. Log minus log plot for proportional hazards checking of year



In this plot, year categories cross each other so the hazard assumptions are not assessed.

Fig.7. Log minus log plot for proportional hazards checking of lactation order



In this plot, lactation categories cross each other so the hazard assumptions are not assessed.

Results of checking assumptions in the Cox proportional hazard regression model using a statistical test (Schoenfeld residual):

The Schoenfeld residuals are defined only at uncensored survival times, for censored observations they are set as missing. The sum of the Schoenfeld residuals for a covariate is zero. Thus, Schoenfeld residuals have a mean of zero. These residuals are not correlated with one another for assessing the proportional hazards assumption.

The residuals are not correlated with each other as in Table (2), so the proportional hazard assumption is assessed.

Table 2. Correlation matrix of eight partial residuals versus ranked survival time

	Rank of days open	PR for age at calving	PR for days in milk	PR for total milk yield	PR for dry period	PR for calving interval	PR for season	PR for year	PR for lactation
Rank of days open									
PR for age at calving	-0.008 0.776								
PR for days in milk	-0.013 0.617	-0.013 0.641							
PR for total milk yield	-0.047 0.078	0.053* 0.049	0.875** 0.000						
PR for dry period	-0.019 0.468	0.533** 0.000	0.060** 0.027	0.152** 0.000					
PR for calving interval	-0.009 0.730	0.578** 0.000	0.133** 0.000	0.266** 0.000	0.750** 0.000				
PR for season	0.017 0.538	-0.094** 0.000	-0.204** 0.000	-0.297** 0.000	-0.096** 0.000	-0.157** 0.000			
PR for year	0.048 0.077	-0.007 0.797	-0.731** 0.000	-0.780** 0.000	-0.091** 0.001	-0.166** 0.000	0.370** 0.000		
PR for lactation	-0.035 0.199	0.660** 0.000	0.027 0.317	0.161** 0.000	0.567** 0.000	0.671** 0.000	-0.139** 0.000	-0.156** 0.000	

** Correlation is significant at the 0.01 level (2-tailed). - * Correlation is significant at the 0.05 level (2-tailed).

-The P-value for testing whether the correlation is zero between the rank of days open and the partial residuals of the variables is the P-value for the statistical test that proportional hazard assumption is assessed or not. Here, there is no correlation assessed and the assumption is assessed.

CONCLUSIONS

This study was conducted on data obtained from a number of Holstein cattle lactation records during the years from 2004 – 2007, at Dina farms which located eighty kilometers north of Cairo.

The data in this study were tested for normality and they were not normally distributed and did not fit a known theoretical distribution.

Based on the K-M survivorship percentiles, median days open for cows were calculated and the median is often better in survival analysis because a small number of individuals with exceptionally long or short lifetimes will cause the mean survival time to be disproportionately large or small.

There was no significant difference between seasons, a highly significant difference between years and lactation order where the P value was less than 0.05.

In Cox's proportional hazard model age at calving, days in milk, season and year have the hazard rate < 1, so the chance of pregnancy is reduced.

Lactation order has the hazard rate > 1, so the chance of pregnancy is increased.

Other variables (total milk yield, dry period and calving interval) did not affect the hazard rate.

From the partial residual test There were no correlation between partial residuals of the variables under study and rank of days open where the P value were larger than 0.05, so the proportional hazard assumption is satisfied.

It is recommended to use survival analysis in veterinary and biological sciences on a large scale. Non parametric survival methods are suitable when the data are not normally distributed. Cox proportional hazard regression model is a good choice to represent many types of data where censoring is present.

REFERENCES

1. ***Zhao GMA (2008):*** Nonparametric and parametric survival analysis of censored data with possible violation of method assumptions. Published Master Thesis. Faculty of The Graduate School at The University of North Carolina at Greensboro.
2. ***Gurcan IS and Akcay A (2007):*** Survival analysis on calving interval and gestation length in Simmental x South Anatolian Red F_1 X B_1 crossbred cows. Ankara Üniv. Vet. Fak. Derg., 54, 219-222.
3. ***Allore HG, Warnick LD, Hertl J and Grohn YT (2001):*** Censoring in survival analysis: A simulation study of the effect of milk yield on conception. Prev. Vet. Med., 46:223-234.
4. ***Ducrocq V (1997):*** Survival analysis, a statistical tool for longevity data. 48th Annual Meeting of the European Association for Animal Production. Vienna, Austria.
5. ***Lee ET and Go OT (1997):*** Survival analysis in public health research. Annu. Rev. Public Health. 18:105–34.
6. ***Dawson B and Trap GR (2004):*** Basic & clinical biostatistics. Boston, McGraw-Hill Professional.
7. ***Bewick V Cheek L and Ball J (2004):*** Statistics review 12: Survival analysis. Critical Care. 8(5): 389–394.
8. ***Nihal A and Haydar D (2013):*** Weighted estimation in Cox regression: An application to breast cancer data. Gazi University Journal of Science. 26(1):63-72.
9. ***SPSS (2007):*** **SPSS/PC+ (2007),** for the PC/XT. SPSS INC.
10. ***Lee ET and Wang JW (2003):*** Statistical methods for survival data analysis. A Wiley-Interscience Publication, John Wiley & Sons, Inc. 3rd Ed., Ch. 1, P. 1, Ch.2, P.8.
11. ***Taufik E and Suriyasataphorn W (2008):*** Survival analysis of the effect of season at calving, lactation number and breeding on days open in dairy cattle. Indonesian Journal of Animal and Veterinary Sciences. Vol.13, No.3.

الملخص العربي

التحليل الإحصائي لبيانات البقاء البيولوجية

*خيري محمد البيومي ، **الحسيني عبدالبر راضي ، *محمود صلاح الطرباني ، *فاطمة دسوقي محمد

* قسم تنمية الثروة الحيوانية - كلية الطب البيطري - جامعة الزقازيق

** قسم الإحصاء التطبيقي والاقتصاد القياسي - معهد الدراسات والبحوث الإحصائية - جامعة القاهرة

أجريت هذه الدراسة علي بيانات خاصة بسجلات الحليب لأبقار الفريزيان النقية التي تم الحصول عليها من مزارع دينا في الفترة ما بين ٢٠٠٤ إلي ٢٠٠٧ ، حيث استهدفت الدراسة تطبيق بعض الأساليب والطرق الإحصائية التي تستخدم لتحليل بيانات البقاء البيولوجية.

تم تحديد الطرق الإحصائية المناسبة لتحليل بيانات البقاء البيولوجية وهي الطرق اللا معلمية مثل:

Kaplan-Meier, Log-rank and Cox's proportional hazard regression model.

تم الاعتماد علي الوسيط لوصف البيانات بدلا من المتوسط وكانت قيمة الوسيط ١٣٤يوم.

اظهرت النتائج عدم وجود فرق معنوي بين الفصول الموسمية.

كما أظهرت النتائج وجود فرق معنوى عالي جدا بين سنوات الولادة عند مستوي معنوية ٥ % و ١ %.

أيضا أظهرت النتائج وجود فرق معنوي عالي جدا بين مواسم الحليب عند مستوي معنوية ٥ % و ١ %.

أظهر نموذج Cox's proportional hazard regression model المتغيرات التي لها تأثير معنوي علي الفترة المفتوحة وبناء عليه لها تأثير علي الحمل وكانت هذه المتغيرات.

- بعض المتغيرات كان لها تأثير واضح علي حدوث الحمل مثل مواسم الحليب المختلفة حيث أدت إلي زيادة فرصة الحمل.
- بعض المتغيرات كان لها تأثير واضح أيضا ولكنها أدت إلي تقليل فرصة حدوث الحمل مثل العمر عند الولادة ، طول فترة الحليب و سنة وموسم الولادة.
- تم اختبار النموذج لتحقيق فروضه الاحصائية باستخدام Schoenfeld residuals test أثبت هذا الإختبار أن كل المتغيرات تحت الدراسة تحقق شروط النموذج حيث أنه وجد انعدام الارتباط بين المتغيرات لذلك يوصي باستخدام هذا النموذج للتعبير عن هذه المتغيرات وتحليلها إحصائيا.
- ومن هذا المنطلق يوصي باستخدام هذه الطرق الإحصائية لتحليل بيانات البقاء في العلوم البيولوجية والبيطرية حيث أن هذه الطرق هي المناسبة لتحليل البيانات التي تحتوي علي قيم مبتورة.
- أيضا يوصي باستخدام Cox's proportional hazard regression model علي نطاق أوسع في العلوم البيولوجية والبيطرية وفي دراسة تأثير بعض المتغيرات الخاصة بمزارع الألبان حيث أنه يعتبر أفضل من غيره من النماذج الإحصائية الأخري لتمثيل البيانات خاصة إذا كانت تحتوي علي قيم مبتورة.