Cox's Model for Prison Partly Interval Censored Data

IOP Conf. Series: Journal of Physics: Conf. Series **1489** (2020) 012032 doi:10.1088/1742-6596/1489/1/012032

Cox's Model for Prison Partly Interval Censored Data

Shaikha Ahmedi[2], Faiz A M Elfaki[2,1], Iing Lukman[3] and N A Kabbashi[4]

[2]Department of Mathematics, Statistics and Physics, College of Art and Sciences, Qatar University, Qatar
[3]Department of Accounting, Faculty of Economics, Universitas Malahayati, Jalan Pramuka 27, Kemiling, Bandar Lampung, Lampung, Indonesia
[4]Department of Biotechnology, Kulliyyah of Engineering, IIUM, Jalan Gombak, P.O. Box 10, 50728 Kuala Lumpur, Malaysia.

Abstract. The term survival analysis has been used in examines and models the time until the events occur. The most common tool for studying the dependency of survival time on predictor variables is Cox model proportional hazards regression model. In this paper we present a simple modification of Cox's proportional hazards model using the partial likelihood principle technique based on Newton Rapson method.
Simulation is conducted based on prison partly interval censored data set with particular sample sizes to evaluate the performance of the proposed model, and it shows that the model is feasible and works well.

1. Introduction

The survival analysis method follows the target subject for a stipulated duration which can be a part of life period or even the entire span of life of the object. The analysis focuses on the time of occurrence of all events pertaining to the target or individual, for analyzing the observed data. The individually recorded data is called the survival data.
The individual can be a person, an animal, or any device like an electrical machine or its component, and the span of data collection can the entire lifetime of the object or a portion thereof. It can also be the curing period of a patient in a hospital or the time period needed for an object to complete a task or the during of an economic cycle.
If the subject is a person, then the event can be the occurrence of a birth, marriage, divorce, disease, death etc. An apt example of a survival data set of a person is the Electronic Health Records (EHRs) which are extensively used in modern state-of-the-art hospitals all over the world to monitor the medical functioning a person.
When the event of interest does not happen or the information pertaining to the survival time of the subject is incomplete, then this kind of observation is called censored. There are three general types of censoring that is: right-censored, left-censored, and one of the most important types of interval censored is partly interval censored data which means that for some of the subjects the event of interest is exactly observed while for others it lies within an interval [9]. In this paper, analysis is conducted for partly interval censored via simulation data set based on prison data set.

[1] Corresponding author. Tel: +974-4403-7564; e-mail: felfaki@qu.edu.qa

Several researcher used statistical methods in the survival data analysis among others, [14] described the survival analysis or failure time analysis as one of the most significant and advanced methods in statistics during the last quarter of the 20th century. [15] described the survival analysis as one of the significant statistical methods as it is involved with the failures of components. [10] described the survival analysis as the procedure to analyze data statistically and the outcome is the time until an event occurs. [11] described that the duration could be compared with lifetime of a marriage; a marriage may end due to annulment, divorce, or death. [4] used survival analysis approach for measuring time until an event occurs and account for teacher's attrition.

Also they are several researcher used survival analysis on criminology among them; [5] analyzed recidivism of three types of reconviction based on competing risks concept. [1] predicted recidivism on a boot camp's graduates using Cox's model. [12] used prisoners in China to studied time to recidivism. [13] obtained the probability of failures for survival analysis based on reincarceration criterion. [2] used hazard function based on semi parametric to studied numbers of crimes averted by incapacitation. [6] used a non parametric based on Kaplan Meier estimation to predicting recidivism on sexual offenders. However, in this research paper we will use Cox model for simulated prison partly interval censored data.

2. The Cox Model

Cox Proportional Hazards Model is one of the most popular models being used extensively in survival analysis. This model envisages the assessment of the significance of various covariates in the survival times of subjects or individuals through the hazard function.

A well-recognized technique for analyzing survival data is Cox model, is based on a modelling approach and aims at exploring the effects of several variables on survival simultaneously. The Cox Model analyses the survival of patients in the clinical trials and the model facilitates to isolate the effects of treatment from the effects of other variables. By theoretical deduction, the model can also be used, if the other variables, which cannot be easily controlled in a clinical trial but affect the patient survival apart from the treatment, are also known.

A Cox Model gives an estimate of the effect of the treatment on survival after adjusting for other variables. This model was introduced by Cox in 1972 [3] for analysis of survival data with and without events. Let t be continuous random variable, $\beta = (\beta_1, ..., \beta_n)'$ be a vector regression parameters, $z = (z_1, ..., z_n)$ be a exploratory variable associated with the individual or covariates and $h_0(t)$ is a baseline hazard. The model can written as;

$$h(t / z) = h_0(t) \exp(z\beta) \tag{1}$$

To inference for the regression model we used standard approach as the numerical methods such as Newton Raphson method or EM algorithm method, ect.

Here, if we observe a subject who failed at time t, then, the contribution to the likelihood is probability density function $f(t, \beta, z)$ at t. The contribution from a subject censored at t is $S(t)$ the survival function beyond t. Thus, partial likelihood based on the data (t_i, δ_i, z_i), $i = 1, 2, ..., n$ is given by [7] and [8] as follows;

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(z_i^T \beta)}{\sum_{k \in R_i(t)} \exp(z_i^T \beta)} \right]^{\delta_i} \tag{2}$$

where δ_i's are event indicator variables: $\delta_i = 1$ if the ith subject fail; $\delta_i = 0$ if the ith subject is censored.

The parameter β is indexes the density function; and z_i are the covariates for the ith subject.

Taking the natural logarithm of equation (2) simplifies the optimization. The log-partial likelihood function is given by [7] as follows:

$$Log(L(\beta)) = \sum_{i=1}^{n} \delta_i \left[z_i^T \beta - \log\left(\sum_{k \in R_i(t)} \exp(z_i^T \beta) \right) \right] \qquad (3)$$

The parameters will be obtained by maximize equation (3) and that will yield the parameters of the model, which are obtained by simultaneously solving the following partial first derivative (the score function $U(B)$) and second derivative (the local information matrix $I(B)$) as;

$$\frac{\partial Log(L(\beta))}{\partial \beta_k}, \frac{\partial^2 Log(L(\beta))}{\partial \beta_k^2}$$

3. Simulation Data

To investigate the effect of Cox model for this prison data set and to compare the variables in the data sets, simulation study was done based on the a prison real data set (which is not addressed here in this paper). We generate the data based on Weibull distribution since we found that the Weibull distribution is appropriated for the real data (The histogram graphics of a real data was found to be related to Weibull distributions curves compared to others distributions). The sample was taken for 1000 times for each variables (age, gender, social status, nationality, previous arrested, and religion).

Table 1 shows the result from our model based on this simulation data sets. The results indicate that the variables age, gender and religion are highly significant compare with other variables such as nationality, previous arrested and social status with respect to the p-value. In addition, the Likelihood Ration Test (LRT) shows a significant different for the age (42.8, with P-value = 5.98e-11), gender (30.5, with P-value=3.27e-08) and religion (52.4 (4.49e-13). It was found that occupancy level is higher among the age 30 and over compared to the younger inmates. This is shown in Figure 1. The study also indicate that males commit more crimes compared to females, as shown in Figure 2. Figure 6 shows the variable religion is highly significant with respect to LRT.

However, Figure 3 shows that there is no significant different between married and single for the social status. Similarly, there is no significant different for Nationality and previous arrested variables as shown in Figure 4 and 5 respectively. (Significance level is 0.05)

Table 1. The result from Cox Model based on simulation data sets.

Variable	Coefficient	Exp (Coef)	SE	P-value	LRT* (P-value)
Age	-0.7820	0.458	0.054	2.2e-16	42.8(5.98e-11)
Gender	-1.281	0.278	0.201	1.7e-10	30.5(3.27e-08)
Social. Status	-0.192	0.826	0.102	0.0610	3.47(0.062500)
Nationality	-0.212	0.809	0.108	0.0500	3.95(0.046700)
Previous	-0.207	0.813	0.122	0.0910	2.97(0.08500)
Religion	-1.133	0.322	0.140	6.7e-16	52.4(4.49e-13)

LRT*: Likelihood Ratio Test

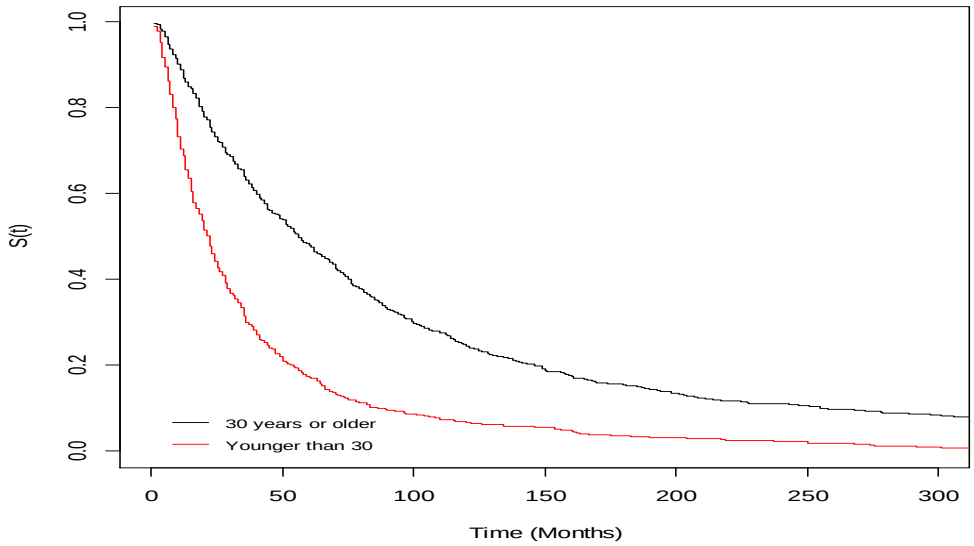

Figure 1: The survival function for age groups (30 years or above and younger than 30 years)

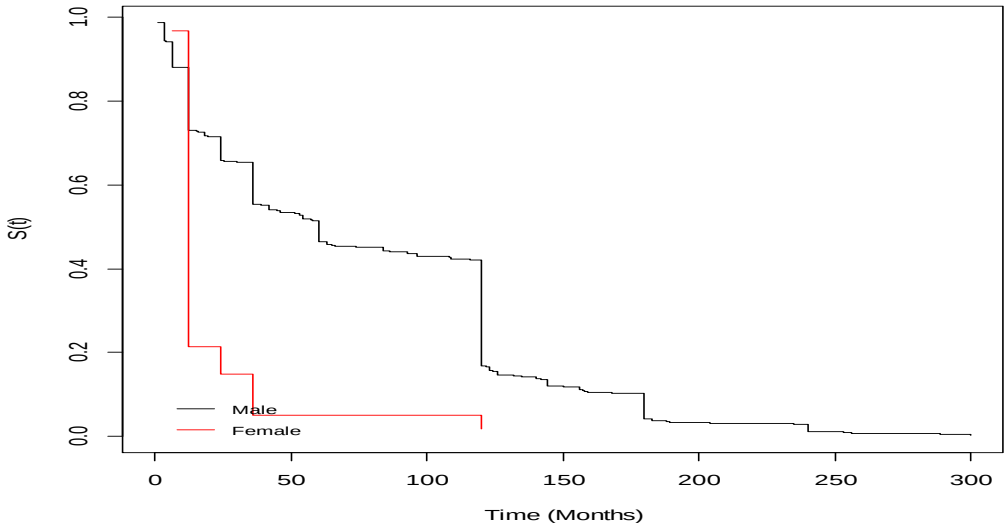

Figure 2: The survival function for gender (male and female)

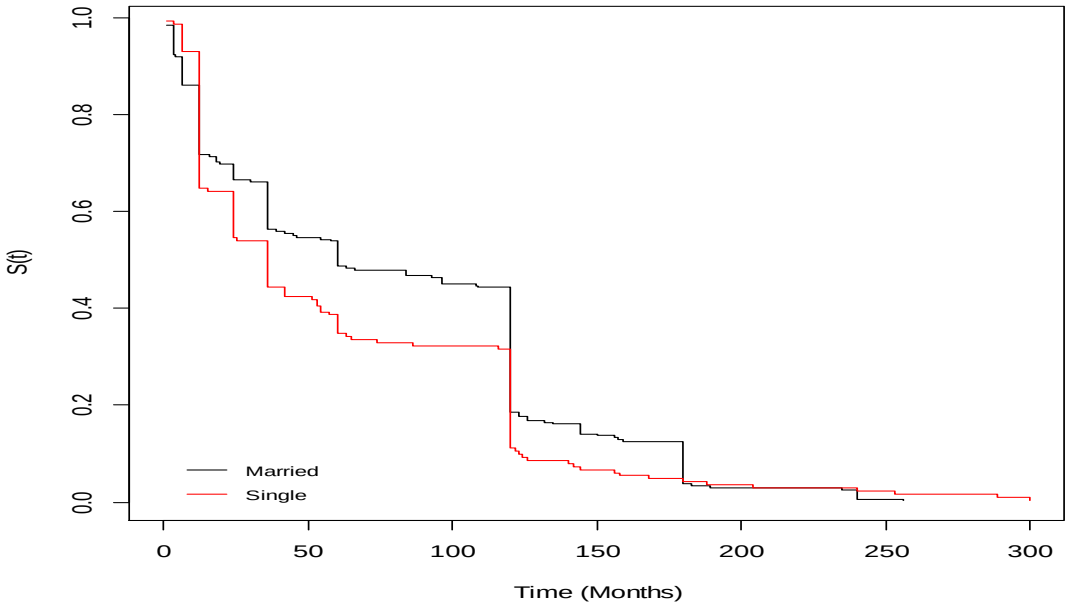

Figure 3: The survival function for Social Status (married and single)

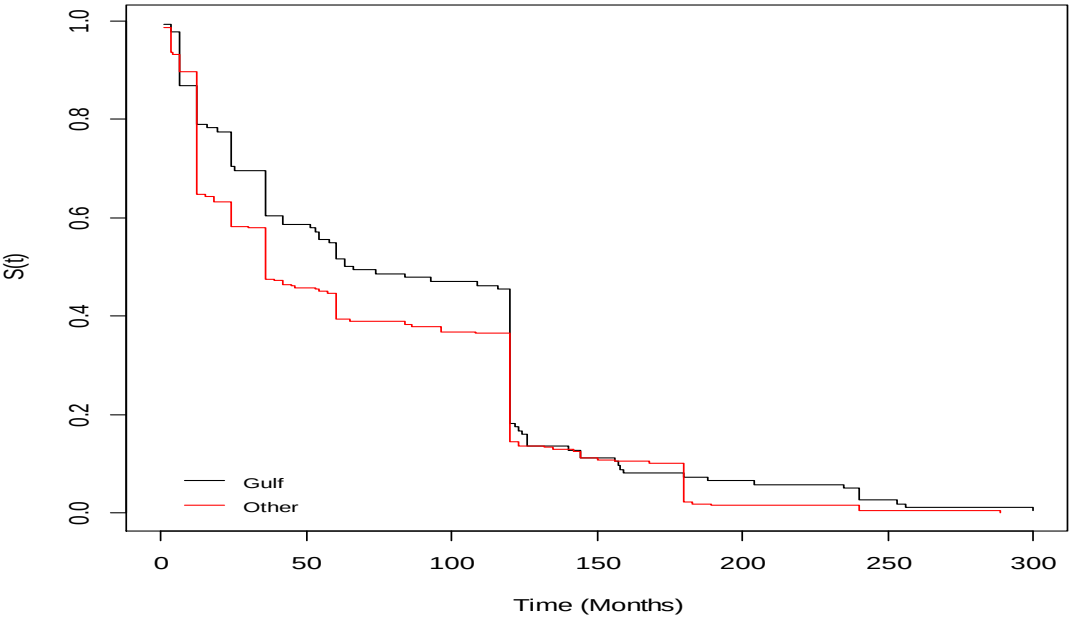

Figure 4: The survival function for Nationality (Gulf and other)

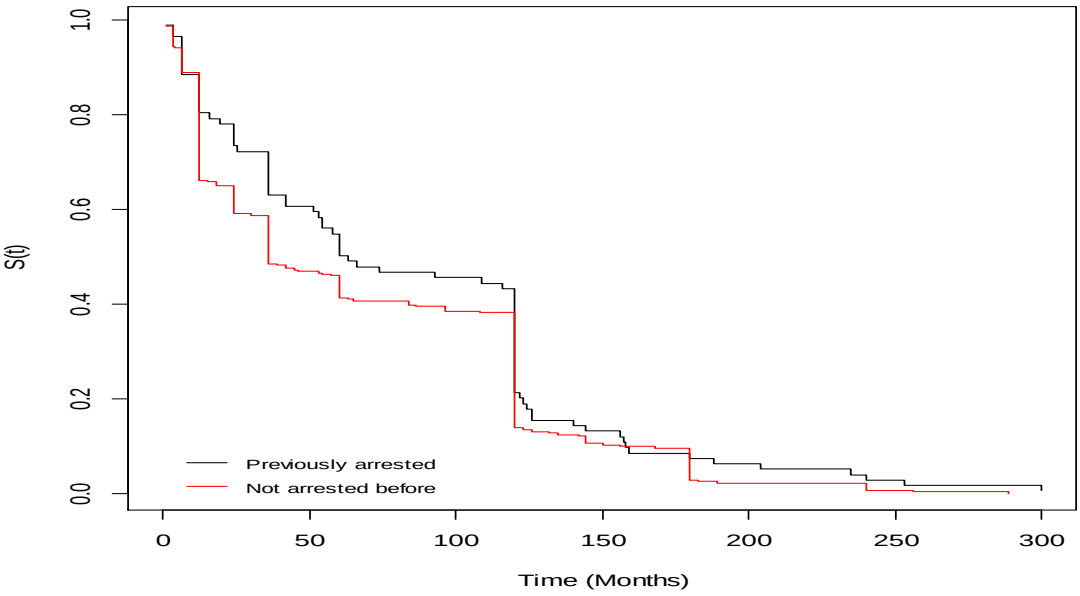

Figure 5: The survival function for arrested (Previous arrested and not arrested before)

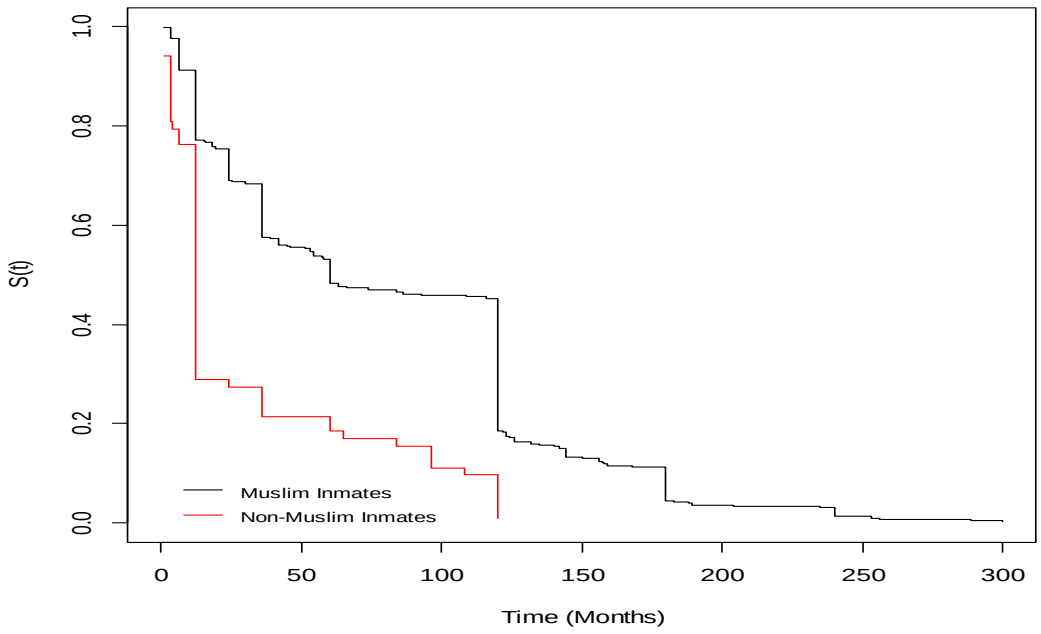

Figure 6: The survival function for Religion (Muslim and non-Muslim)

4. Conclusion

We have proposed a Cox's proportional hazards regression model for prison partly-interval censored data. Simulations studies based on the prison real data set (which is not addressed here in this paper) were used to investigate the performance of our model. We generate the data based on Weibull distribution for six variables, which are; age, gender, social status, nationality, religion and previous arrested.

The preliminary results obtained from the simulated data demonstrate that the developed model based on partly-interval censored data performs well and it is applicable. More improvements in the proposed model will be introduced later in another paper with illustration using a real data set.

References

[1] Benda B B 2005 Gender Differences in Life-Course Theory of Recidivism: A Survival Analysis *International Journal of Offender Therapy and Comparative Criminology* **49** 325-342.

[2] Bhati A S 2007 Estimating The Number of Crimes Averted by Incapacitation: An Information Theoretic Approach *Journal of Quantitative Criminology* **23** 355—375.

[3] Cox D R 1972 Regression models and life tables *J Roy Statist Soc* B **34** 187–220.

[4] Eagle and Barnes 2014 Survival Analysis on Duration Data in Intelligent Tutors *Intelligent Tutoring Systems* **8474** pp 178-187.

[5] Escarela G Francis, B., and Soothill, K. 2000 Competing Risks, Persistence, and Desistance in Analyzing Recidivism *Journal of Quantitative Criminology* **16** 385-414.

[6] Hill A, Habermann N, Klusmann D, Berner W, and Briken P 2008 Criminal Recidivism in Sexual Homicide Perpetrators *International Journal of Offender Therapy and Comparative Criminology* **52** 5 - 20.

[7] Kalbfleisch J D, and Lawless J F 1988 Estimation of reliability in field-performance studies *Technimetrics* **30** 365-388.

[8] Kalbfleisch J D and Prentice R L 1980 The Statistical Analysis of Failure Time Data (New York: Wiley)

[9] Kim J S 2003 Maximum Likelihood Estimation for the Proportional Hazards Model with Party Interval-Censored Data *J R. Statist. Soc., Series* B **65** 489-502.

[10] Kleinbaum and Klein 2005 *Survival analysis: A self-learning text* (NewYork: Springer)

[11] Lawless 2003 *Statistical Models and Methods for Lifetime Data* (Canada:John Wiley & Sons)

[12] Liu J 2005 Predicting Recidivism in A Communitarian Society: China *International Journal of Offender Therapy and Comparative Criminology* **49** 392-409.

[13] Loza W, Neo L H, Shahinfar S and Loza-Fanous A 2005 Cross-Validation of The Self-Appraisal Questionnaire: A Tool for Assessing Violent And Nonviolent Recidivism with Female Offenders *International Journal of Offender Therapy and Comparative Criminology* **49** 547-560

[14] Sam and Krings 2008 Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management *IEEE Aerospace Conference.*

[15] Singh and Totawattage 2013 The Statistical Analysis of Interval-Censored Failure Time Data with Applications *Journal of Statistics* **3** pp 155-166