

# Survival Analysis in Modeling the Birth Interval of the First Child in Indonesia

Rahmat Hidayat, Hadi Sumarno, Endar H. Nugrahani

Department of Applied Mathematics, Faculty of Mathematics and Natural Science, Bogor Agricultural University, Bogor, Indonesia

Email: [dayatmath@gmail.com](mailto:dayatmath@gmail.com)

Received 21 February 2014; revised 21 March 2014; accepted 28 March 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

First birth interval is one of the examples of survival data. One of the characteristics of survival data is its observation period that is fully unobservable or censored. Analyzing the censored data using ordinary methods will lead to bias, so that reducing such bias required a certain method called survival analysis. There are two methods used in survival analysis that are parametric and non-parametric method. The objective of this paper is to determine the appropriate method for modeling the birth of the first child. The exponential model with the inclusion of covariates is used as parametric method, considering that the newly married couples tend to have desire for having baby as soon as possible and such desire will be weakened by increasing age of marriage. The data that will be analyzed were taken from the Indonesia Demographic and Health Survey (IDHS) 2012. The result of data analysis shows that the birth of the first child data is not exponentially distributed thus the *Cox proportional hazard* method is used. Because of the suspicion that disproportional covariate exists, then the proportional *hazard* test is conducted to show that the covariate of age is not proportional, the generalized *Cox proportional* method is used, namely *Cox extended* that allows the inclusion of disproportional covariates. The result of analysis using *Cox extended* model indicates that the factors affecting the birth of the first child in Indonesia are the area of residence, educational history and its age.

## Keywords

Analysis Survival; Censored; Covariate; *Cox Proportional Hazard*; *Cox Extended*

---

## 1. Introduction

In demography, there are three things which take effect, namely mortality, migration, and fertility. The birth interval of the first child can be used as one of indicators of fertility. The birth interval of the first child is defined

as difference between the age of marriage and the age of birth of the first child. In fact, the length of birth interval of the first child of every married woman is not same. According to existing research, the birth interval of the first child is determined by many kinds of social and culture even physiology factors. According to [1], there are several factors that affect the birth interval of the first child, which are the area of residence, educational level, age, the knowledge of contraception and employment status. Most of married women have given birth to their first child a couple months right after the marriage, so that the obtained data are the complete one, but the others who do not have children yet are classified as the censored data. Thus the birth interval of the first child is an example of the survival data. The survival data are the data that indicate the period of time since the initial observation to a happened event. The characteristic of the survival data is its survival time, which usually could not be fully observed (censored) [2]. If all the expected events have happened, and are able to be fully observed, some analysis methods can be done; unfortunately, the *survival* data are censored [3]. Analyzing the survival data using the ordinary method would be inappropriate because of causing bias [4]. A certain method to analyze it was needed for reducing such bias. This certain method is called *survival* analysis.

## 2. Definitions Associated with Survival Analysis

Definition 1: Survival time measures the period of time from the initial event to happened event such as failure, death, respond, symptoms, and etc. [5].

Definition 2: Survival function is a function indicates the probability of an individual who survives until or more than  $t$  time (experiencing the event after  $t$  time) [6]. Consider random variable  $T$ , then the *survival* function define as,  $S(t) = P(T > t)$ . Consider  $f$  as the probability density function, then the *survival* function also defines as complement of cumulative function  $F$  as,

$$S(t) = \int_t^{\infty} f(x) dx = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad [7]$$

Definition 3: *Hazard* function is a function indicates the probability of an individual of having risk or experiencing event such as failure or die at  $t$  time in the condition that this individual survived to  $t$  time, the function given by:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \quad [8]$$

From those definitions above, the relation between the *survival* function and *hazard* function are obtained. Using the definition of the conditional probability, obtained:

$$P(t \leq T \leq t + \delta t | T \geq t) = \frac{P(t \leq T < t + \delta t)}{P(T \geq t)} = \frac{F(t + \delta t) - F(t)}{S(t)}$$

$$h(t) = \lim_{\delta t \rightarrow \infty} \left( \frac{F(t + \delta t) - F(t)}{\delta t} \right) \frac{1}{S(t)} h(t) = \frac{f(t)}{S(t)}$$

$$\text{Since } f(t) = -\frac{dS(t)}{dt}$$

$$\text{Then } h(t) = -\frac{d(\ln S(t))}{dt}$$

$$\ln S(t) = -\int_0^t h(u) du$$

$$\ln S(t) = -H(t)$$

$$S(t) = \exp(-H(t))$$

## 3. Model

In this research, there are two survival methods that will be used, *i.e.* exponential parametric and non-parametric

method.

### 3.1. Parametric Method

Survival time can be analyzed by using the *accelerated failure time* (AFT) model. In the survival time, this model assumes that the logarithm relation of survival time  $T$  and its covariates are linear and can be written as

$$\log T = a_0 + \sum_{j=1}^p a_j x_j + \sigma \varepsilon$$

where  $x_j, j=1,2,\dots,p$  indicate its covariates,  $a_j, j=0,1,\dots,p$  coefficient of  $\sigma$  indicates the scale parameter and  $\varepsilon$  indicates the error.

[9] states that for inclusion of covariates in the exponential distribution, we use the equation above and choose  $\sigma=1$  to obtain

$$\log T_i = a_0 + \sum_{j=1}^p a_j x_{ji} + \varepsilon = \mu_i + \varepsilon_i$$

where  $\mu_i = a_0 + \sum_{j=1}^p a_j x_{ji}$ .  $T$  is the exponential distribution with *hazard* function, probability density function, and *survival* function respectively

$$h(t, \lambda_i) = \lambda_i = \exp \left[ - \left( a_0 + \sum_{j=1}^p a_j x_{ji} \right) \right] = \exp(-\mu_i)$$

$$h(t, \lambda_i) = \lambda_i \exp(-\lambda_i t)$$

$$S(t, \lambda_i) = \exp(-\lambda_i t)$$

### 3.2. Non-Parametric Method

#### 3.2.1. Kaplan-Meier Model

In the Kaplan-Meier method, assuming the distribution of the data is discrete. According to [10], Kaplan-Meier is a method used to compare the *survival* time of two covariate groups. The advantage of this method is that the non-parametric method does not require the knowledge of a particular distribution [11].

This method is appropriate because the data used are the individual data, but still appropriate for small, medium and large data sizes. Suppose the time of the birth of the first child denotes by  $r$  and the number of women who are married denotes by  $n$ , where  $r \leq n$ . The probability of the birth of the first child in every interval  $j$  estimated by  $\frac{d_j}{n_j}$ , and its *survival* probability estimated by  $\hat{p}_j = \left( \frac{n_j - d_j}{n_j} \right)$ . The estimator for *survival* function of Kaplan-Meier method is given by,

$$\hat{S}(t) = \hat{p}_1 \times \hat{p}_2 \times \hat{p}_3 \times \dots \times \hat{p}_k = \prod_{j=1}^k \hat{p}_j = \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right)$$

where  $t_k \leq t < t_{k+1}, k=1,2,3,\dots,m$ .  $\hat{S}(t) = 1$  for  $t \leq t_1$ .

#### 3.2.2. Cox Proportional Hazard and Non-Proportional Hazard Model

*Cox proportional hazard* is a model usually used as multivariate approach to analyze the data [12]. The characteristic of *Cox proportional hazard* model indicates that every different individual has proportional *hazard* function, that is,  $\left[ \frac{h(t|x_1)}{h(t|x_2)} \right]$ , the ratio of the *hazard* function of two individuals with the inclusion of covariates  $x_1 = (x_{11}, x_{21}, \dots, x_{p1})'$  and  $x_2 = (x_{12}, x_{22}, \dots, x_{p2})'$  are constant. This means that the ratio of failure risk of two individuals is the same and does not depend on how long they survived.

[13] explains that the general form of *Cox proportional hazard* model is:

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}).$$

where  $x$  denotes covariate, but he did not make any assumptions about the form of  $h_0(t)$  itself, which is called the baseline of *hazard* function, because it is the value of *hazard* function when  $x=0$ .

Sometimes the time-dependent covariate was found, so it is not met with the proportional assumption, then the form above was developed into *Cox extended* model:

$$h_i(t, x(t)) = h_0(t) \exp(\beta_1 x_{1i} + \dots + \beta_{p1} x_{p1i} + \delta_1 x_{1i}(t) + \dots + \delta_{p2} x_{p2i}(t))$$

For checking the proportional assumption of covariate, the *Cox extended* model can be written as:

$$h(t, x(t)) = h_0(t) \exp\left[\sum_{j=1}^p \beta_j x_j + \sum_{j=1}^p \delta_j x_j g_j(t)\right]$$

where  $g_j(t)$  denotes function of time and it is important to determine the proper form of  $g_j(t)$ . The following possibilities of a function  $g_j(t)$  according to [14] are:

- 1)  $g_j(t) = 0$  is the simplest form by resulting the *Coxproportional hazard* model.
- 2)  $g_j(t) = t$ . If the test result  $\delta_j$  is significant then the *Cox extended* model is better than the *Cox proportional hazard* model, thus the *ratio* of *hazard* function is a function of time.
- 3)  $g_j(t) = \log(t)$ .
- 4)  $g_j(t)$  *heavyside function*. When this function is used then obtained the constant of *hazard* ratio for different time intervals.

#### 4. Parameter Estimation

In estimating parameter  $\beta_1, \beta_2, \dots, \beta_p$ , Cox using the *maximum likelihood estimation* method (the maximum *likelihood* estimator) by only considering the individual probability who are experiencing the event called *partial likelihood* [14]. Estimation  $\beta_j$  using *partial likelihood* means maximizing the *partial likelihood* function. *Partial likelihood* function is the joint probability *survival* function of uncensored data formed by the function of unknown value of parameter. [7] stated that parameter estimation  $\beta$  can be proved by taking *survival* individual cases such as the death event. Suppose there are  $n$  individuals with  $r$  individuals who are having death then there exist  $(n-r)$  individuals censored. Assumed that there is only an individual who died at a certain time of death (there is no *ties*). Again suppose that  $t_{(1)} < t_{(2)} < t_{(3)} < \dots < t_{(r)}$  is the ordered and uncensored *survival* time. The probability of the  $i$ -th individual death at time  $t_{(j)}$  in terms of  $t_{(j)}$  as the only exact time of death of  $t_{(1)}, t_{(2)}, t_{(3)}, \dots, t_{(r)}$  and the covariate for died individual at time  $t_{(j)}$  is  $x_{(j)}$  denote as:

$$P(\text{person with covariate } x_{(j)} \text{ died at time } t_{(j)} | \text{single death at time } t_{(j)}) \\ = \frac{P(\text{person with covariate } x_{(j)} \text{ died at time } t_{(j)})}{P(\text{single death at time } t_{(j)})}$$

The numerator indicates the  $i$ -th individual death risk at time  $t_{(j)}$ , denotes as  $h_i(t_{(j)})$ , while the denominator indicates the number of death risk at time  $t_{(j)}$  for all individual who had death risk at  $t_{(j)}$  or its sum  $h_i(t_{(j)})$  in  $R(t_{(j)})$ , where  $R(t_{(j)})$  is the set of individual who had risk of death at time  $t_{(j)}$ , that is individuals who lived and uncensored just before  $t_{(j)}$  so that  $R(t_{(j)})$  called the *risk set*. Suppose A is the  $i$ -th individual event with covariate  $x_{(j)}$  died at time  $t_{(j)}$  and B is the single death event at time  $t_{(j)}$ . The above equation becomes

$$P(A|B) = \frac{h_i(t_{(j)})}{\sum_{l \in R(t_{(j)})} h_l(t_{(j)})}$$

By substituting equation in *Cox proportional hazard* model obtained

$$P(A|B) = \frac{h_0(t) \exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} h_0(t) \exp(\beta' x_l)} = \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)}$$

Thus the *likelihood* function of the conditional probability above is

$$L(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}' \mathbf{x}_l)}$$

where  $\mathbf{x}_{(j)}$  is a covariate vector for individual who died at time  $t_{(j)}$ . The amount of  $\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}' \mathbf{x}_l)$  represent the sum of  $\exp(\boldsymbol{\beta}' \mathbf{x}_l)$  for every element of individual  $R(t_{(j)})$ . Multiplication in the *likelihood* function only for uncensored individual. The censored individual excluding in the numerator but in the denominator is the sum of  $\exp(\boldsymbol{\beta}' \mathbf{x}_l)$  for every  $l$  element of  $R(t_{(j)})$ .

Suppose that the data consist of  $n$  observation of *survival* time  $t_1, t_2, t_3, \dots, t_n$  and  $\delta_i$  denotes the indicator of event by valued

$$\delta_i = \begin{cases} 0, & \textit{ith individual right censored} \\ 1, & \textit{others.} \end{cases}$$

Then the equation of *likelihood* function can be written as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right\}^{\delta_i}$$

If this equation turned to be logarithm then obtained

$$\begin{aligned} \log L(\boldsymbol{\beta}) &= \log \left( \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right\}^{\delta_i} \right) = \sum_{i=1}^n \left( \log \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right\}^{\delta_i} \right) \\ &= \sum_{i=1}^n \delta_i \left( \log \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right) = \sum_{i=1}^n \delta_i \left\{ \boldsymbol{\beta}' \mathbf{x}_i - \log \sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \right\}. \end{aligned}$$

Parameter estimator  $\boldsymbol{\beta}$  can be obtained by maximizing the log function  $L(\boldsymbol{\beta})$ , so that the solution be

$$\frac{d \log L(\boldsymbol{\beta})}{d \boldsymbol{\beta}} = 0 \Leftrightarrow \sum_{i=1}^n \delta_i \left\{ \mathbf{x}_i - \frac{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l) \mathbf{x}_l}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right\} = 0.$$

The solution of such equation is hardly solved analytically but more easily solved using numeric method.

## 5. Data

The data used in this research taken from the result of Indonesia Demographic and Health Survey in 2002. The samples used are the data of two provinces, that are West Papua and Special Area of Yogyakarta as representation of high and low fertility level area. The data limited to the birth interval of the first child, of first time married woman.

## 6. Operational Definition

The dependent variable used in this research is the birth interval of the first child of a first time married woman. Whereas the independent variables that estimated to affect the birth interval of the first child are:

1) *Area of residence*, grouped in the smallest administrative unit of area, that are urban and countryside area. The area of residence divided into two categories, that are city = 1 and village = 2.

2) *Education*, schools are the formal schools start from the primary, junior, and high school, including the equated education. Those who never sign into formal education or ever been in primary school but never getting a passing mark are classified as not finished the primary school. The highest education divided into four categories, that are not finished the primary school = 0, finished the primary school = 1, finished the junior high school = 2, and finished the senior high school or higher = 3.

3) *Employment status*, working is the activity of doing the job with purpose of obtaining or earning income or profit for at least one hour per week continuously and uninterrupted (including as unpaid family workers who

helped in the business/economic activity). Employment status are categorized into unemployment = 0 and employment = 1.

- 4) *Knowledge about contraception*, divided by two categories, unknown = 0 and known = 1.
- 5) *Age of mother*, age of mother/first time married woman expressed in years.

## 7. Results

The Kolmogorov-Smirnov goodness of fit test can be applied to test whether a sample theoretically follows a population distribution. Based on the calculating results obtained that the value of  $D = 0.258$  and the critical value of table  $D^*$  for significance level 0.05 is 0.04025. Since the calculating value of  $D$  is greater than the critical value of table  $D$  then reject  $H_0$ , which means that the data is not exponentially distributed. Furthermore, since the data of birth interval of the first child is not exponentially distributed then the exponential method is not used for analyzing the factors that supposed dominantly affecting the birth interval of the first child.

As an illustration, the birth interval data of the first child will be analyzed using the Kaplan-Meier method. Its respond variable is the time of marriage to the time of having the first child and the area of residence is the variable that affect the *survival* level (0 = village, 1 = city).

The *survival* analysis results using the Kaplan-Meier method for independent area of residence of variable can be shown in graphic as follows.

From **Figure 1** shown that the survived individual who lived in the villages are different with those who lived in the cities. Next the hypothesis test will be conducted to see whether this difference significantly affect or just coincidence, by the following hypothesis,

$$H_0 : S_1(t) = S_2(t)$$

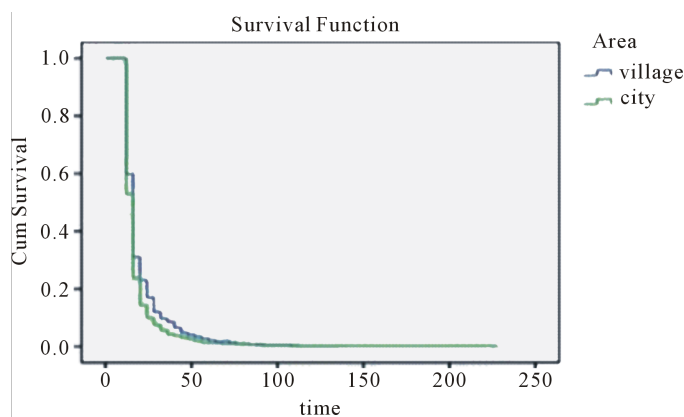
$$H_1 : S_1(t) \neq S_2(t)$$

If the log rank  $W_L > \chi^2_{(1)\alpha}$  is the statistic test value then the rejection area is  $H_0$ . The log rank test results based on the area of residence covariate status are presented in the **Table 1**.

By using the significance level 0.05 of chi-square test with degree of freedom 1 obtained  $W$  value significant enough to reject  $H_0$ . So, it can be concluded that there is significantly difference among the *survival* level of the birth of the first child, who lived in the village and in the city.

If the *survival* data that will be compared are comes from more than two individual groups, e.g. we are interested in seeing the characteristic difference of the area of residence, education level, age, knowledge about contraception and so forth, then the Kaplan-Meier method will be impractical. This caused of in the Kaplan-Meier method, every two of population groups must be tested separately, so if there are several groups then repeated testing must be conducted. If the respondents have several characteristics, then the *Cox proportional hazard* method able to explain the influence of such characteristics to the respond variables simultaneously.

**Table 2** shows that the explanatory variables significantly affected to the interval birth of the first child are the area of residence, education 1, education 2, education 3, and age.



**Figure 1.** Survival function of Kaplan-Meier method for independent area of residence variables.

**Table 1.** The log rank test result to see the difference of the *survival* level of the birth of the first child based on the area of residence covariate.

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	9136	1	0.003

**Table 2.** Correlation and *p*-value of explanatory variables.

Explanatory variables	Correlation	<i>p</i> -Value
Area of residence	0.05438	0.1489
Education 1	0.02557	0.4976
Education 2	-0.05312	0.1586
Education 3	0.03078	0.4141
Knowledge of contraception	0.01169	0.7565
Age	-0.10459	0.0054*
Job	-0.04565	0.2257

Next the estimation to the proportional assumption of every covariate was done. In this paper, the only method used to check the covariate that does not met the proportional assumption in the birth interval of the first child is Schoenfeld residuals method. The results are presented in **Table 2**.

The covariate of *p*-value of individual aged less than 0.05 means there is correlation between such covariate to the time rating until that individual had her first child, so that the age covariate does not met the proportional assumption. Thus the data re-modeled using the *Cox extended* model. Before the analysis is done, conducted the test previously to see more appropriate form of  $g_j(t)$ . This test uses the AIC criteria. The AIC is a measure for selecting the best regression model, introduced by Hirotugu Akaike in 1973. The AIC method is based on the *maximum likelihood estimation* [15], with the equation as follows

$$AIC = -2 \log \hat{L} + \alpha q$$

where  $\hat{L}$  indicates the *likelihood* function,  $q$  indicates the sum of parameter  $\beta$ , and  $\alpha$  indicates the specified constant. The value of  $\alpha$  that often used is between 2 and 6. Based on the AIC method, the best regression model is the model that has the smallest value of AIC [9].

The value of AIC formed in  $g_1(t) = \log t$  is the smallest among the two other models (**Table 3**). Thus this is the best model.

Next the test was done to see significantly affected covariate to the respond variable. The results are presented in **Table 4**.

In **Table 4** seen that the explanatory variables that significantly affect to the birth interval of the first child are the area of residence, education 1, education 2, education 3, and age.

## 8. Interpretation

The value of *hazard* ratio allows us to compare among multiple groups in the *survival* analysis [16]. **Table 5** shows that the area of residence, education 1, education 2, education 3, and age variables are significantly affect to the individual risk of having first child. From the value of *hazard* ratio seen that individual who lived in the city had 0.720 times lower risk of having their first child than who lived in village one, and the educational history of individual who finished their primary, junior, senior high school or higher will increase the risk of having their first child respectively 1.708, 2.648, 4.361 times of individual who had lower educational history or not finished their primary school at all. Every additional age of one year will reduce the risk of having first child as 4.5%.

**Table 4** shows that the area of residence, education 1, education 2, education 3, and age variables are the significantly affected to the individual risk of having her first child. From the value of *hazard* ratio seen that individual who lived in the city had 0.719 times lower risk of having their first child than who lived in village one,

**Table 3.** The value of AIC model.

No.	Model Type	The Value of AIC
1	<i>Cox proportional hazard</i>	10464.278
2	<i>Cox extended <math>g_1(t) = t</math></i>	10465.014
3	$g_1(t) = \log t$	10451.802

**Table 4.** Parameter estimator, *p*-value, and the *hazard ratio* by using the *cox extended* model.

Explanatory variables	Parameter estimators	Standard error	Chi-Square	<i>p</i> -Value	<i>Hazard ratio</i>
Area of residence	-0.32996	0.05896	31.3204	<0.0001	0.719
Education 1	0.54800	0.20091	7.4396	0.0064	1.730
Education 2	0.98723	0.20295	23.6637	<0.0001	2.684
Education 3	1.44341	0.22071	42.7709	<0.0001	4.235
Knowledge of contraception	0.23638	1.01985	0.0537	0.8167	1.267
Age	0.06833	0.03079	4.9249	0.0265	1.071
Job	-0.07083	0.01017	1.6276	0.2020	0.932
Age $t$	-0.03861	0.01017	14.3992	0.0001	0.962

**Table 5.** Parameter estimator, *p*-value, and *hazard ratio* using *cox proportional hazard* model.

Explanatory variables	Parameter estimator	Standard error	Chi-Square	<i>p</i> -Value	<i>Hazard ratio</i>
Area of residence	-0.32835	0.05896	31.0109	<0.0001	0.720
Education 1	0.53510	0.20082	7.1002	0.0077	1.708
Education 2	0.97377	0.20284	23.0461	<0.0001	2.648
Education 3	1.47264	0.22033	44.6750	<0.0001	4.361
Knowledge of contraception	0.21333	1.01079	0.0438	0.8343	1.238
Age	-0.04625	0.00712	42.2332	<0.0001	0.955
Job	-0.06678	0.05547	1.4495	0.2286	0.935

and the educational history of individual who finished her primary, junior, senior high school or higher will increase the risk of having their first child respectively 1.730, 2.648, 4.235 times of individual who had lower educational history or not finished their primary school at all. Every additional age of one year will reduce the risk of having first child as 3.0166%.

## 9. Conclusions

The result of distribution tests by using the Kolmogorov-Smirnov test indicates that the data do not have a particular distribution, so the *Cox proportional hazard* model of non-parametric method is a more appropriate method for modeling the birth interval data of the first child. However, after testing the assumption of *proportional hazard* model, it turned out that one of the covariates did not meet the proportional assumption, so that the data re-modeled by using generalized *Cox proportional hazard* model called *Cox extended model*.

Although there are no differences among the explanatory variables that significantly affect the using of *Cox proportional* as *Cox extended* model, but the interpretation of both models is still different and based on the conducted test. The *Cox extended* model is the best model in modeling the birth data of the first child in Indonesia.



## Acknowledgements

The first, authors would like to thankful to Allah SWT, my Parents (Syahrudin and Salmah K), lecturer, and my friends for support for this paper.

## References

- [1] Ibrohim, J. (1994) Analisis selang kelahiran anak di Jawa Barat. Jurusan Statistika Fakultas Matematikadan Ilmu Pengetahuan Alam [Tesis]: Bogor (ID): Program Pascasarjana, Institut Pertanian Bogor.
- [2] Epstein, B. and Sobel, M. (1953) Life Testing. *Journal of the American Statistical Association*, **48**, 486-502. <http://dx.doi.org/10.1080/01621459.1953.10483488>
- [3] Clark, T.G., Bradburn, M.J., Love, S.B. and Altman, D.G. (2003) Survival Analysis Part I: Basic Concepts and First Analysis. *British Journal of Cancer*, **89**, 232-238. <http://dx.doi.org/10.1038/sj.bjc.6601118>
- [4] Widyaningsih, Y. (2006) Penerap analisis regresi logistic dan analisis survival pada masa laktasi wanita Indonesia [Tesis]. Bogor (ID): Program Pascasarjana, Institut Pertanian, Bogor.
- [5] Lee, E.T. (1992) Statistical Methods for Survival Data Analysis. 2nd Edition, Wiley Interscience Publication, New York.
- [6] Banerjee, T. (2007) Bayesian Analysis of Generalized Odds-Rate Hazards Models for Survival Data. *Lifetime Data Analysis*, **13**, 241-260. <http://dx.doi.org/10.1007/s10985-007-9035-3>
- [7] Collet, D. (1994) Modelling Survival Data in Medical Research. 2nd Edition, Chapman & Hall/CRC, London. <http://dx.doi.org/10.1007/978-1-4899-3115-3>
- [8] Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data. University Press, Cambridge.
- [9] Love, C., Altman, D.G. and Bradburn, M. (2003) Multivariate Data Analysis. *British Journal of Cancer*, **89**, 437-443.
- [10] Hoon, T.S. (2008) Using Kaplan-Meier and Cox Regression in Survival Analysis. *Journal ESTEEM*, **4**, 3-14.
- [11] Kaplan, E.L. and Meier, P. (1958) Non Proportional Estimation from Incomplete Observation. *Journal of the America Statistical Association*, **53**, 457-481. <http://dx.doi.org/10.1080/01621459.1958.10501452>
- [12] Bradburn, M.J., Clark, T.G. and Love, S.B. (2003) Multivariate Data Analysis, an Introduction to Concepts and Methods. *British Journal of Cancer*, **89**, 431-436. <http://dx.doi.org/10.1038/sj.bjc.6601119>
- [13] Cox, D.R. (1972) Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society: Series B*, **34**, 187-220.
- [14] Kleinbaum, D.G. and Klein, M. (2012) Survival Analysis a Self-Learning Text. 3rd Edition, Springer, New York.
- [15] Fathurahman, M. (2009) Pemilihan model regresi terbaik menggunakan metode Akaike's Information Criterion an Schwarz Information Criterion. *Jurnal Informatika Mulawarman*, **4**, 37-41.
- [16] Maetani, S. and Gamel, J. (2013) Parametric Cure Model versus Proportional Hazard Model in Breast Cancer and Other Malignancies. *Advances in Breast Cancer Research*, **2**, 119-125. <http://dx.doi.org/10.4236/abcr.2013.24020>