Analisis *Survival* Data Kejadian Bersama dengan Pendekatan Efron *Partial Likelihood* (Studi Kasus: Lama Masa Studi Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2011)

Survival Analysis Ties Event Data Using Efron Partial Likelihood Approach (Case Study: Duration Study of Student of Faculty of Mathematics and Natural Science Mulawarman University 2011)

Santi Prabawati[1], Yuki Novia Nasution[2], dan Sri Wahyuningsih[3]
[1,3]Laboratorium Statistika Terapan Jurusan Matematika FMIPA Universitas Mulawarman
[2]Laboratorium Matematika Komputasi Jurusan Matematika FMIPA Universitas Mulawarman
[1]E-mail: santistat13@gmail.com

Abstract

Survival analysis is a statistical procedure used to analyze data related to survival time, from the defined time origin until the occurrence of certain events. In the survival analysis, sometimes ties are found, in which two or more individuals experience the same event at the same time. There are three widely used methods to treat ties in survival analysis, that is Exact method, Breslow approach, and Efron approach. Efron's approach has a simple, fast, and accurate calculation especially when the data contains many ties. The purpose of this study is to find out the Cox proportional hazard data ties model using Efron partial likelihood approach and to know the variables that affect the graduation time of student of Faculty of Mathematics and Natural Sciences of Mulawarman University class of 2011 that graduated until February 28, 2017. The variables are Gender, home area, funding sources, and GPA. Based on the results of the analysis that has been done with the help of software R, it is obtained that the variables that have significant effect are gender and GPA. For the gender variables it was concluded that female students had a chance of 1,362 times to graduate faster than male students. While for the GPA variable it is concluded that each addition of GPA of 0.1, then the student's chance to graduate faster will increase by 1,225 times.

Keywords: Cox Proportional Hazard, Efron Partial Likelihood, Survival, Ties

Pendahuluan

Analisis *survival* atau dikenal sebagai analisis ketahanan hidup merupakan prosedur statistika yang digunakan untuk menganalisis data berupa waktu antar kejadian. Analisis ini digunakan ketika kasus berkaitan dengan waktu atau lama waktu hingga terjadi peristiwa tertentu (Kleinbaum dan Klein, 2005). Salah satu yang membedakan analisis *survival* dengan prosedur analisis statistika yang lain yaitu terdapat konsep penyensoran. Data tersensor merupakan data yang tidak bisa diamati secara utuh karena adanya individu yang keluar dari pengamatan namun belum mengalami *event* atau pada akhir pengamatan individu tersebut belum mengalami *event* tertentu.

Salah satu tujuan analisis *survival* adalah untuk mengetahui hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival*. Hubungan tersebut dapat dimodelkan dengan model regresi Cox *proportional hazard*. Model tersebut memiliki asumsi bahwa fungsi *hazard* dari individu yang berbeda adalah proporsional atau rasio fungsi *hazard* dari dua individu yang berbeda adalah konstan (Lee dan Wang, 2003).

Dalam analisis *survival* terkadang ditemukan adanya kejadian bersama atau yang sering disebut *ties*. *Ties* adalah keadaan di mana terdapat dua individu atau lebih yang mengalami kejadian pada waktu yang bersamaan. Terdapat tiga metode yang biasa digunakan untuk mengatasi kejadian bersama dalam analisis *survival* yaitu metode *Exact*, pendekatan Breslow, dan pendekatan Efron. Pendekatan Efron memiliki perhitungan yang sederhana, cepat, dan akurat terutama ketika data mengandung banyak *ties*.

Penelitian sebelumnya yang terkait dengan analisis *survival* pada kejadian bersama dilakukan oleh Rahmadeni dan Ranti (2016) yang membandingkan model Cox menggunakan estimasi parameter Efron *partial likelihood* dan Breslow *partial likelihood* pada data pasien diabetes. Hasil penelitian menunjukkan bahwa metode Efron *partial likelihood* lebih baik dalam mengestimasi nilai parameter regresi dibanding metode Breslow *partial likelihood* di mana metode Efron mempunyai nilai AIC yang lebih kecil dibanding dengan model Cox yang menggunakan estimasi parameter Breslow *partial likelihood*. Penelitian lain dilakukan oleh Vitriana dan Kusumawati (2016) yang membahas model Cox *extended* dengan metode Breslow pada kejadian bersama yang diterapkan pada data individu berhenti bekerja.

Pada penelitian ini, analisis *survival* data kejadian bersama dengan pendekatan Efron *Partial Likelihood* akan diaplikasikan pada data lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 sampai dengan 28 Februari 2017.Tujuan dari penelitian ini adalah untuk mengetahui model Cox *proportional hazard* pada kejadian bersama dengan pendekatan Efron *partial likelihood* dan mengetahui variabel-variabel yang mempengaruhi lama waktu kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 yang lulus sampai dengan 28 Februari 2017.

Analisis *Survival*

Analisis *survival* atau dikenal sebagai analisis ketahanan hidup merupakan prosedur statistika yang digunakan untuk menganalisis data berupa waktu antar kejadian. Analisis ini digunakan ketika kasus berkaitan dengan waktu atau lama waktu hingga terjadi peristiwa tertentu (Kleinbaum dan Klein, 2005).

Waktu *Survival*

Waktu *survival* dapat didefinisikan sebagai waktu dari awal pengamatan hingga terjadinya peristiwa gagal. Waktu awal (*time origin* atau *start-point*) yaitu waktu pada saat kejadian awal, seperti tanggal perawatan pasien, sedangkan waktu kegagalan (*failure time* atau *end-point*) yaitu waktu pada saat terjadinya kejadian akhir seperti meninggal atau sembuh seorang pasien (Collet, 2003).

Fungsi Kepadatan Peluang

Fungsi kepadatan peluang adalah peluang suatu individu mengalami kejadian dalam interval waktu dari t sampai $t + \Delta t$. Fungsi kepadatan peluang dinotasikan dengan $f(t)$ dan dirumuskan dengan

$$f(t) = \lim_{\Delta t \to 0}\left[\frac{P\left(t \le T < \left(t + \Delta t\right)\right)}{\Delta t}\right] \tag{1}$$

Misalkan T adalah variabel random non negatif pada interval $[0, \infty)$ yang menunjukkan waktu hidup pada suatu populasi dan $f(t)$ merupakan fungsi kepadatan peluang dari t maka fungsi distribusi kumulatif $F(t)$ adalah (Lawless, 2003):

$$F(t) = P(T \le t) == \int_0^t f(x)\,dx \tag{2}$$

sehingga dari Persamaan (2) diperoleh

$$f(t) = \frac{d\left(F(t)\right)}{dt} = F'(t) \tag{3}$$

Fungsi *Survival*

Fungsi *survival* dinotasikan dengan $S(t)$, yaitu peluang suatu individu bertahan atau tidak mengalami kejadian lebih lama dari waktu t yang dirumuskan dengan (Lee dan Wang, 2003):

$$\begin{aligned} S(t) &= P(T > t) = 1 - P(T \le t) \\ &= 1 - F(t) \end{aligned} \tag{4}$$

Fungsi *Hazard*

Fungsi *hazard* merupakan tingkat risiko terjadinya suatu *event*. Menurut Lee dan Wang (2003), fungsi *hazard* $h(t)$ dari waktu *survival* T memberikan tingkat kegagalan bersyarat yang didefinisikan sebagai peluang kegagalan selama interval waktu yang sangat kecil, dengan asumsi bahwa individu tersebut bertahan hingga awal interval, atau sebagai batas peluang bahwa individu gagal dalam interval yang sangat kecil, $t + \Delta t$, mengingat individu tersebut bertahan hingga waktu t. Fungsi *hazard* dapat dituliskan dengan persamaan sebagai berikut:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0}\frac{P(t \le T < \left(t + \Delta t\right) | T > t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \end{aligned} \tag{5}$$

Menurut Lee dan Wang (2003), fungsi kumulatif *hazard* didefinisikan sebagai

$$H(t) = \int_0^t h(x)\,dx \tag{6}$$

maka hubungan antara fungsi kumulatif *hazard* dengan fungsi *survival* yaitu

$$S(t) = \exp\left[-H(t)\right] \tag{7}$$

Estimasi Fungsi *Survival* Kaplan-Meier

Estimasi Kaplan-Meier disebut juga estimasi *product limit*. Estimasi fungsi *survival* ini pertama kali dikemukakan oleh Kaplan dan Meier pada tahun 1958. Misalkan T variabel random non negatif pada interval $[t_i, t_{i+1})$. Estimasi fungsi *survival* didefinisikan sebagai berikut (Hosmer dan Lemeshow, 2008):

$$\hat{S}(t) = \prod_{t_i < t}\frac{n_i - d_i}{n_i} \tag{8}$$

di mana:

$\hat{S}(t)$= estimasi fungsi *survival*
n_i = jumlah individu yang berisiko mengalami *event* pada saat t_i
d_i = jumlah individu yang mengalami *event* pada saat t_i

Kejadian Bersama

Dalam analisis survival terkadang ditemukan adanya kejadian bersama atau yang sering disebut *ties*. *Ties* adalah keadaan di mana terdapat dua individu atau lebih yang mengalami kejadian pada waktu yang bersamaan. Jika suatu data terdapat *ties*, maka akan menimbulkan permasalahan dalam membentuk *partial likelihood* yaitu saat menentukkan anggota dari himpunan risikonya. Terdapat tiga metode yang biasa digunakan untuk mengatasi kejadian bersama dalam analisis *survival* yaitu metode *Exact*, pendekatan Breslow, dan pendekatan Efron. Pendekatan Efron memiliki perhitungan yang sederhana, cepat, dan akurat terutama ketika data mengandung banyak *ties* (Allison, 2010).

Asumsi *Proportional Hazard*

Kleinbaum dan Klein (2005) menyebutkan bahwa ada beberapa pendekatan yang dapat dilakukan untuk pemeriksaan asumsi *proportional hazard* yaitu dengan pendekatan grafik $\ln[-\ln S(t)]$ dan uji *goodness of fit* dengan residual Schoenfeld. Menurut Collet (2003), ada tiga pilihan untuk mengatasi Cox *nonproportional hazard* yaitu dengan mengeluarkan variabel bebas yang tidak memenuhi asumsi dari model, menggunakan model Cox stratifikasi, dan dengan perluasan model Cox.

1. Pendekatan Grafik $\ln[-\ln S(t)]$

Dalam menggunakan grafik $\ln[-\ln S(t)]$, data dikelompokkan sesuai dengan kategori pada masing-masing variabel bebas. Apabila grafik $\ln[-\ln S(t)]$ menunjukkan kurva yang sejajar, maka asumsi *proportional hazard* terpenuhi. (Collet, 2003).

2. Uji *Goodness of Fit*

Pengujian asumsi dengan pendekatan *goodness of fit* merupakan variasi dari tes Schoenfeld, yaitu menggunakan residual Schoenfeld. Residual Schoenfeld untuk individu yang mengalami *event* pada saat t_i pada variabel bebas ke-j adalah sebagai berikut:

$$R_{ij} = \delta_i \left(x_{ij} - \frac{\sum_{l \in R(t_i)} x_{lj} \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{x}_l\right)}{\sum_{l \in R(t_i)} \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{x}_l\right)} \right), j = 1,2,\ldots,p; i = 1,2,\ldots,n \tag{9}$$

di mana δ_i adalah indikator penyensoran, $\delta_i = 1$ untuk data tidak tersensor, $\delta_i = 0$ untuk data tersensor dan $\hat{\boldsymbol{\beta}}$ merupakan estimator *partial likelihood* maksimum dari $\boldsymbol{\beta}$. Karena $\hat{\boldsymbol{\beta}}$ merupakan solusi dari persamaan turunan pertama fungsi *log likelihood*, maka jumlah residual Schoenfeld adalah nol atau dengan kata lain residual Schoenfeld mempunyai rata-rata nol. Apabila jumlah sampel besar, nilai harapan dari R_{ij} adalah nol, sehingga residual Schoenfeld tidak berkorelasi dengan yang lainnya (Lee dan Wang, 2003).

Adapun langkah-langkah pengujian asumsi *proportional hazard* menggunakan residual Schoenfeld adalah sebagai berikut:

1. Membangun model Cox *proportional hazard* dan mencari taksiran residual Schoenfeld untuk setiap variabel prediktor.
2. Mengurutkan waktu *survival* mulai dari individu yang mengalami *event* pertama kali.
3. Menguji korelasi antara residual Schoenfeld dengan peringkat waktu *survival*.

Hipotesis yang diuji adalah sebagai berikut:

H_0: Tidak terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival*

H_1: Terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival*

Statistik uji yang digunakan adalah:

$$z = \frac{r}{\frac{1}{\sqrt{n-1}}} \tag{10}$$

di mana:

$$r = \frac{\sum_{i=1}^{r}\left(R_{ij} - \bar{R}_{ij}\right)\left(RT_i - \overline{RT}_i\right)}{\sqrt{\sum_{i=1}^{r}\left(R_{ij} - \bar{R}_{ij}\right)}\sqrt{\sum_{i=1}^{n}\left(RT_i - \overline{RT}_i\right)}}$$

R_{ij} = residual Schoenfeld untuk individu yang mengalami *event* pada saat t_i untuk variabel bebas ke-j

$\bar{R}_{ij}$ = rata-rata residual Schoenfeld untuk individu yang mengalami *event* pada saat t_i untuk variabel bebas ke-j

RT_i = peringkat waktu *survival* untuk waktu kejadian ke-i

$\overline{RT}_i$ = rata-rata peringkat waktu *survival* untuk waktu kejadian ke-i

Apabila nilai $|z| > z_{\alpha/2}$ atau $p - value \leq \alpha$ maka dapat disimpulkan bahwa terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival* sehingga asumsi *proportional hazard* tidak terpenuhi.

Model Regresi Cox *Proportional Hazard*

Model regresi Cox *proportional hazard* adalah salah satu cara yang digunakan untuk mengetahui hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival*. Model tersebut memiliki asumsi bahwa fungsi *hazard* dari individu yang berbeda adalah proporsional atau rasio fungsi *hazard* dari dua individu yang berbeda adalah konstan (Lee dan Wang, 2003). Secara umum, bentuk dari model regresi Cox *proportional hazard* adalah:

$$h(t, \boldsymbol{x}) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p\right) \tag{11}$$

dengan:

$h_0(t)$ = fungsi *baseline hazard*

$\beta_1, \beta_2, \ldots, \beta_p$ = parameter regresi

$x_1, x_2, \ldots, x_p$ = variabel bebas

Estimasi Parameter Model Regresi Cox *Proportional Hazard*

Estimasi parameter β_j dengan $j = 1,2, \ldots, p$ pada model regresi Cox *proportional hazard* dapat dilakukan salah satunya menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE). Metode MPLE hanya mempertimbangkan peluang individu yang mengalami kejadian saja yang disebut *partial likelihood*. Fungsi *partial likelihood* merupakan fungsi peluang bersama dari data *survival* tidak tersensor yang biasa dinotasikan dengan $L(\boldsymbol{\beta})$, di mana parameter-parameter $\boldsymbol{\beta}$ tidak diketahui nilainya (Kleinbaum dan Klein, 2005).

Misalkan terdapat data untuk n individu yang terdiri dari r waktu kejadian yang tidak tersensor dan $n-r$ individu yang tersensor kanan. Apabila diurutkan, urutannya menjadi $t_1 < t_2 < \cdots < t_i < \cdots < t_r$, dengan t_i merupakan urutan kejadian ke-i. Diasumsikan hanya terdapat satu individu yang meninggal pada tiap waktu kegagalan sehingga tidak terjadi *ties* pada data. Apabila individu-individu yang berisiko mengalami kematian pada waktu t_i dinotasikan dengan l dan $R(t_i)$ merupakan himpunan dari individu-individu yang berisiko pada waktu t_i, maka

$$P(A|B) = \frac{h_i(t_i)}{\sum_{l\in R(t_i)} h_l(t_i)} \quad (12)$$

Berdasarkan hasil peluang bersyarat pada Persamaan (12), maka diperoleh fungsi *partial likelihood* sebagai berikut:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p}\beta_j x_{ij})}{\sum_{l\in R(t_i)} \exp(\sum_{j=1}^{p}\beta_j x_{lj})} \quad (13)$$

Dari Persamaan (13) diperoleh fungsi *log partial likelihood* sebagai berikut (Klein dan Moeschberger, 2003):

$$\ln L(\boldsymbol{\beta}) = \ln\left[\prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p}\beta_j x_{ij})}{\sum_{l\in R(t_i)} \exp(\sum_{j=1}^{p}\beta_j x_{lj})}\right]$$
$$= \sum_{i=1}^{r}\left[\left(\sum_{j=1}^{p}\beta_j x_{ij}\right) - \left(\ln \sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)\right)\right] \quad (14)$$

Estimasi Parameter Model Regresi Cox *Proportional Hazard* dengan Pendekatan Efron *Partial Likelihood*

Fungsi *partial likelihood* dengan pendekatan Efron adalah sebagai berikut:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp\left(\sum_{l\in D(t_i)}\sum_{j=1}^{p}\beta_j x_{lj}\right)}{\prod_{k=1}^{d_i}\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)\right]} \quad (15)$$

di mana:

$D(t_i)$ = himpunan individu yang mengalami *event* pada saat t_i

$R(t_i)$ = himpunan individu yang berisiko mengalami *event* pada saat t_i

k = individu yang mengalami *ties*

d_i = banyaknya *ties* yang teramati pada saat t_i

Dari Persamaan (15) diperoleh fungsi *log partial likelihood* sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \ln\left[\prod_{i=1}^{r} \frac{\exp\left(\sum_{l\in D(t_i)}\sum_{j=1}^{p}\beta_j x_{lj}\right)}{\prod_{k=1}^{d_i}\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)\right]}\right]$$
$$= \sum_{i=1}^{r}\sum_{l\in D(t_i)}\sum_{j=1}^{p}\beta_j x_{lj} -$$
$$\sum_{i=1}^{r}\sum_{k=1}^{d_i}\left[\ln\left(\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)\right)\right] \quad (16)$$

Pengujian Signifikansi Parameter

Pengujian signifikansi parameter regresi Cox *proportional hazard* dilakukan secara serentak menggunakan uji *partial likelihood ratio* dan secara individu menggunakan uji Wald.

1. Uji *Likelihood Ratio*

Uji *likelihood ratio* dinotasikan dengan G, digunakan untuk menguji pengaruh variabel bebas dalam model regresi secara serentak. Hipotesis dari pengujian ini adalah:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1: minimal ada satu $\beta_j \neq 0$, dengan $j = 1, 2, \ldots, p$

Rumus umum statistik uji G adalah sebagai berikut (Hosmer dan Lemeshow, 2008):

$$G = 2\left[\ln L\left(\hat{\beta}_j\right) - \ln L(0)\right]$$
$$= -2\left[\ln L(0) - \ln L\left(\hat{\beta}_j\right)\right] \quad (17)$$

di mana:

$\ln L(0)$ = *log partial likelihood* dari model regresi Cox *proportional hazard* tanpa variabel bebas.

$\ln L(\hat{\beta}_j)$ = *log partial likelihood* dari model regresi Cox *proportional hazard* yang terdiri dari p variabel bebas.

Statistik uji G mengikuti distribusi *chi-square* dengan derajat bebas p. Apabila nilai $G \geq \chi^2_{(\alpha;p)}$ atau $p-value \leq \alpha$ maka dapat disimpulkan bahwa minimal ada satu variabel bebas yang berpengaruh signifikan terhadap waktu *survival*.

2. Uji Wald

Uji Wald digunakan untuk menguji pengaruh parameter (β_j) secara individu, dinotasikan dengan W^2. Hipotesis dalam pengujian ini adalah:

$H_0: \beta_j = 0$, dengan $j = 1, 2, \ldots, p$

$H_1: \beta_j \neq 0$, dengan $j = 1, 2, \ldots, p$

Rumus umum statistik uji Wald adalah sebagai berikut:

$$W^2 = \left(\frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)}\right)^2 \quad (18)$$

di mana:

W^2 = statistik uji Wald

$\hat{\beta}_j$ = koefisien variabel bebas ke-j

$SE(\hat{\beta}_j)$ = *standar error* koefisien variabel bebas ke-j

Statistik uji Wald mengikuti distribusi *chi-square* dengan derajat bebas 1. Apabila nilai $W^2 \geq \chi^2_{(\alpha;1)}$ atau $p-value \leq \alpha$ maka dapat disimpulkan bahwa variabel bebas ke-j berpengaruh signifikan terhadap waktu *survival* (Collet, 2003).

Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 yang lulus sampai dengan 28 Februari 2017, dengan variabel terikatnya yaitu lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 terhitung mulai tanggal 1 Juli 2011 hingga mencapai kelulusan yang dinyatakan dalam satuan bulan dan berada dalam batas periode penelitian dengan ketentuan:

$$\text{Status} = \begin{cases} 1, \text{Jika mahasiswa menempuh studi kurang} \\ \quad \text{dari atau sama dengan 10 semester} \\ 0, \text{Jika mahasiswa menempuh studi melebihi} \\ \quad \text{10 semester} \end{cases}$$

Adapun variabel bebas yang digunakan adalah sebagai berikut:

a. Jenis kelamin (x_1), dengan kategori sebagai berikut:

$$x_1 = \begin{cases} 1, \text{ untuk jenis kelamin perempuan} \\ 0, \text{ untuk jenis kelamin laki-laki} \end{cases}$$

b. Daerah asal (x_2), dengan kategori sebagai berikut:

$$x_2 = \begin{cases} 1, \text{ untuk daerah asal dari luar Samarinda} \\ 0, \text{ untuk daerah asal dariSamarinda} \end{cases}$$

c. Sumber pendanaan (x_3), dengan kategori sebagai berikut:

$$x_3 = \begin{cases} 1, \text{ untuk sumber pendanaan dari} \\ \quad \text{beasiswa bidik misi} \\ 0, \text{ untuk sumber pendanaan bukan dari} \\ \quad \text{beasiswa bidik misi} \end{cases}$$

d. IPK (x_4)

Estimasi Parameter Model Cox dengan Pendekatan Efron *Partial Likelihood*

Langkah pertama yang akan dilakukan adalah membentuk fungsi *partial likelihood* dengan pendekatan Efron yang didefinisikan sebagai berikut:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{166} \frac{\exp\left(\sum_{l\in D(t_i)} \sum_{j=1}^{4} \beta_j x_{lj}\right)}{\prod_{k=1}^{d_i}\left[\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right]} \quad (19)$$

Selanjutnya membentuk fungsi *log partial likelihood* dari $L(\boldsymbol{\beta})$ sehingga diperoleh sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \ln\left[\prod_{i=1}^{166} \frac{\exp\left(\sum_{l\in D(t_i)} \sum_{j=1}^{4} \beta_j x_{lj}\right)}{\prod_{k=1}^{d_i}\left[\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right]}\right]$$
$$= \sum_{i=1}^{166}\sum_{l\in D(t_i)}\sum_{j=1}^{4}\beta_j x_{lj} -$$
$$\sum_{i=1}^{166}\sum_{k=1}^{d_i}\left[\ln\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)\right] \quad (20)$$

Setelah membentuk fungsi *log partial likelihood*, selanjutnya menentukan turunan pertama dari $\ln L(\boldsymbol{\beta})$ terhadap β_j yaitu sebagai berikut:

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\left(\sum_{i=1}^{166}\left[\left(\sum_{l\in D(t_i)}\sum_{j=1}^{4}\beta_j x_{lj}\right) - \sum_{k=1}^{d_i}\left[\ln\left(\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)\right]\right]\right)$$
$$= \sum_{i=1}^{166}\sum_{l\in D(t_i)}\sum_{j=1}^{4} x_{lj} - \sum_{i=1}^{166}\sum_{k=1}^{d_i}\left(\frac{\sum_{l\in R(t_i)}\sum_{j=1}^{4} x_{lj}\left(\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)}\right) \quad (21)$$

Langkah selanjutnya adalah menentukan turunan *partial* kedua dari fungsi *log partial likelihood* yang diperoleh dari turunan *partial* pertama dari fungsi *log partial likelihood* sebelumnya, sehingga diperoleh sebagai berikut:

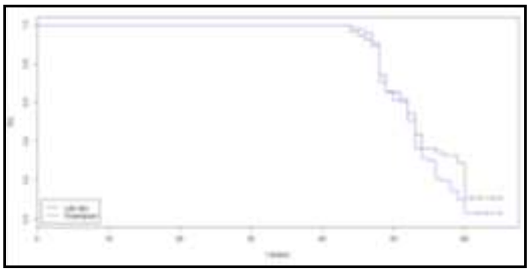
$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial^2 \beta_j^2} = \frac{\partial^2}{\partial^2 \beta_j^2}\left(\sum_{i=1}^{166}\left[\left(\sum_{l\in D(t_i)}\sum_{j=1}^{4}\beta_j x_{lj}\right) - \sum_{k=1}^{d_i}\left[\ln\left(\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)\right]\right]\right)$$
$$= \sum_{i=1}^{166}\sum_{k=1}^{d_i}\left[\frac{\left[\sum_{l\in R(t_i)}\sum_{j=1}^{4} x_{lj}\left(\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)\right]^2}{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right]^2} - \frac{\left[\sum_{l\in R(t_i)}\sum_{j=1}^{4} x_{lj}^2\left(\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right)\right]}{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right) - \frac{k-1}{d_i}\sum_{l\in D(t_i)}\exp\left(\sum_{j=1}^{4}\beta_j x_{lj}\right)\right]}\right] \quad (22)$$

Berdasarkan hasil turunan *partial* kedua yang diperoleh dengan pendekatan Efron *partial likelihood*, dapat dilihat bahwa fungsi tidak dapat diselesaikan secara analitik untuk mendapatkan hasil estimasi masing-masing parameter. Pada kasus seperti ini, maka metode iterasi numerik dapat digunakan sebagai salah satu cara untuk membantu menyelesaikan proses estimasi parameter tersebut. Salah satu metode yang sering digunakan yaitu metode Newton-Raphson. Iterasi Newton-Raphson untuk mengestimasi parameter model Cox dilakukan dengan bantuan *software* R.

Karakteristik Fungsi *Survival* dan Fungsi Kumulatif *Hazard* dengan Metode Kaplan-Meier

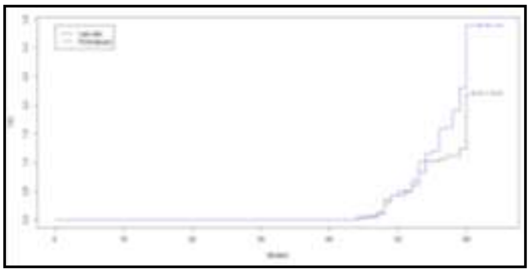
Karakteristik fungsi *survival* dan fungsi kumulatif *hazard* mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman digambarkan dengan grafik berdasarkan variabel-variabel yang diduga mempengaruhi lama masa studi yaitu jenis kelamin, daerah asal, dan sumber pendanaan. Estimasi fungsi *survival* dan fungsi kumulatif *hazard* dengan metode Kaplan-Meier dilakukan dengan bantuan *software* R berdasarkan Persamaan (8).

Karakteristik berdasarkan variabel jenis kelamin sebagai berikut:



Gambar 5. Fungsi *survival* berdasarkan jenis kelamin

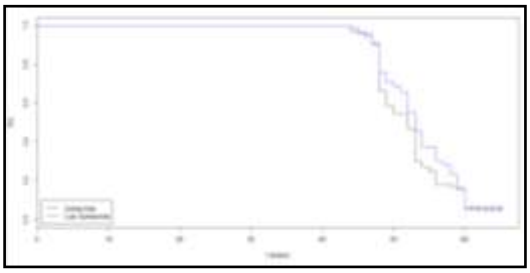
Berdasarkan Gambar 5 dapat dilihat bahwa grafik fungsi *survival* untuk jenis kelamin laki-laki dan perempuan saling berhimpit dan masing-masing turun secara lambat. Pada bulan ke-44 hingga bulan ke-47, peluang mahasiswa yang berjenis kelamin laki-laki untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa yang berjenis kelamin perempuan. Sedangkan pada bulan ke-48 hingga bulan ke-65, peluang mahasiswa yang berjenis kelamin perempuan untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa yang berjenis kelamin laki-laki.



Gambar 6. Fungsi kumulatif *hazard* berdasarkan jenis kelamin

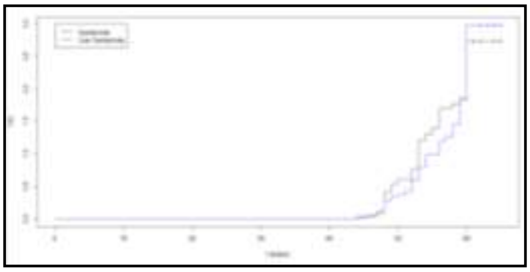
Berdasarkan Gambar 6 dapat dilihat bahwa grafik fungsi kumulatif *hazard* untuk jenis kelamin laki-laki dan perempuan saling berhimpit dan masing-masing naik secara lambat. Pada bulan ke-44 hingga bulan ke-47, kesempatan mahasiswa yang berjenis kelamin laki-laki untuk lulus lebih tinggi dibandingkan mahasiswa yang berjenis kelamin perempuan. Sedangkan pada bulan ke-48 hingga bulan ke-65, kesempatan mahasiswa yang berjenis kelamin perempuan untuk lulus lebih tinggi dibandingkan mahasiswa yang berjenis kelamin laki-laki.

Karakteristik berdasarkan variabel daerah asal sebagai berikut:



Gambar 7. Fungsi *survival* berdasarkan daerah asal

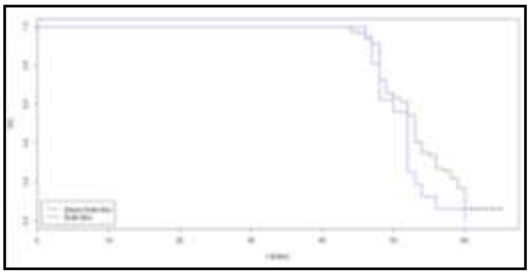
Berdasarkan Gambar 7 dapat dilihat bahwa grafik fungsi *survival* untuk asal daerah Samarinda dan luar Samarinda saling berhimpit dan masing-masing turun secara lambat. Pada bulan ke-44 hingga bulan ke-47, peluang mahasiswa yang berasal dari luar Samarinda untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa yang berasal dari Samarinda. Sedangkan pada bulan ke-48 hingga bulan ke-59, peluang mahasiswa yang berasal dari Samarinda untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa yang berasal dari luar Samarinda. Dan pada bulan ke-60 hingga bulan ke-65, peluang mahasiswa yang berasal dari luar Samarinda untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa yang berasal dari Samarinda.



Gambar 8. Fungsi kumulatif *hazard* berdasarkan daerah asal

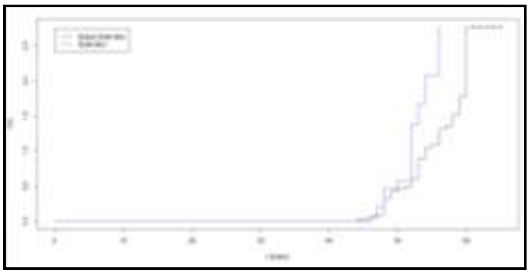
Berdasarkan Gambar 8 dapat dilihat bahwa grafik fungsi kumulatif *hazard* untuk asal daerah Samarinda dan luar Samarinda saling berhimpit dan masing-masing naik secara lambat. Pada bulan ke-44 hingga bulan ke-47, kesempatan mahasiswa yang berasal dari luar Samarinda untuk lulus lebih tinggi dibandingkan mahasiswa yang berasal dari Samarinda. Sedangkan pada bulan ke-48 hingga bulan ke-59, kesempatan mahasiswa yang berasal dari Samarinda untuk lulus lebih tinggi dibandingkan mahasiswa yang berasal dari luar Samarinda. Dan pada bulan ke-60 hingga bulan ke-65, kesempatan mahasiswa yang berasal dari luar Samarinda untuk lulus lebih tinggi dibandingkan mahasiswa yang berasal dari Samarinda.

Karakteristik berdasarkan variabel sumber pendanaan sebagai berikut:



Gambar 9. Fungsi *survival* berdasarkan sumber pendanaan

Berdasarkan Gambar 9 dapat dilihat bahwa grafik fungsi *survival* untuk sumber pendanaan bukan bidik misi dan bidik misi saling berhimpit dan masing-masing turun secara lambat. Pada bulan ke-44 hingga bulan ke-46, peluang mahasiswa yang bukan penerima beasiswa bidik misi untuk tetap menjalani studi lebih rendah dibandingkan mahasiswa penerima beasiswa bidik misi. Sedangkan pada bulan ke-47 hingga bulan ke-60, peluang mahasiswa penerima beasiswa bidik misi untuk tetap menjalani studi cepat lebih rendah dibandingkan mahasiswa yang bukan penerima beasiswa bidik misi.



Gambar 10. Fungsi kumulatif *hazard* berdasarkan sumber pendanaan

Berdasarkan Gambar 10 dapat dilihat bahwa grafik fungsi kumulatif *hazard* untuk sumber pendanaan bukan bidik misi dan bidik misi saling berhimpit dan masing-masing naik secara lambat. Pada bulan ke-44 hingga bulan ke-46, kesempatan mahasiswa yang bukan penerima beasiswa bidik misi untuk lulus lebih tinggi dibandingkan mahasiswa penerima beasiswa bidik misi. Sedangkan pada bulan ke-47 hingga bulan ke-60, kesempatan mahasiswa penerima beasiswa bidik misi untuk lulus lebih tinggi dibandingkan mahasiswa yang bukan penerima beasiswa bidik misi.

Pengujian Asumsi *Proportional Hazard*

Pengujian asumsi *proportional hazard* dilakukan dengan uji *goodness of fit* menggunakan residual Schoenfeld pada persamaan (9). Adapun hasil uji *goodness of fit* dengan bantuan *software* R terdapat pada Tabel 3.

Hipotesis yang digunakan untuk pengujian asumsi *proportional hazard* dengan *goodness of fit* adalah sebagai berikut:

H_0: Tidak terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival*

H_1: Terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival*

Berdasarkan Tabel 3, diperoleh *p-value* lebih besar dari $\alpha = 0{,}1$ untuk semua variabel sehingga maka dari hasil pengujian H_0 ditolak. Dengan tingkat kepercayaan 90% dapat disimpulkan bahwa tidak terdapat korelasi antara residual Schoenfeld dengan peringkat waktu *survival* sehingga semua variabel kategorik telah memenuhi asumsi *proportional hazard*. Untuk itu selanjutnya, dapat digunakan model regresi Cox *proportional hazard*.

Tabel 1. Hasil Pengujian Asumsi *Proportional Hazard* dengan *Goodness of Fit*

Variabel	*p-value*	Keputusan
Jenis Kelamin	0,154	H_0 gagal ditolak
Daerah Asal	0,132	H_0 gagal ditolak
Sumber Pendanaan	0,986	H_0 gagal ditolak

Analisis Regresi Cox *Proportional Hazard*

Setelah melakukan pengujian asumsi *proportional hazard*, tahap selanjutnya adalah pembentukan model regresi Cox *proportional hazard* untuk mengetahui variabel bebas yang berpengaruh terhadap lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam angkatan 2011. Hasil estimasi parameter model regresi Cox *proportional hazard* dengan pendekatan Efron *partial likelihood* dengan bantuan *software* R sebagai berikut:

Tabel 2. Hasil Estimasi Parameter Model Cox untuk Semua Variabel

Variabel	β	*p-value*	Keputusan
Jenis Kelamin	0,302	0,0829	H_0 ditolak
Daerah Asal	-0,053	0,7394	H_0 gagal ditolak
Sumber Pendanaan	0,086	0,7618	H_0 gagal ditolak
IPK	1,976	$7{,}92 \times 10^{-6}$	H_0 ditolak
Likelihood Ratio Test		$7{,}208 \times 10^{-6}$	H_0 ditolak

Selanjutnya, parameter yang diperoleh akan diuji secara serentak dan secara individu untuk mengetahui variabel yang masuk ke dalam model. Adapun hipotesis untuk pengujian secara serentak adalah sebagai berikut:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1: minimal ada satu $\beta_j \neq 0$, dengan $j = 1, 2, 3, 4$

Berdasarkan Tabel 4 diperoleh *p-value* sebesar $7{,}208 \times 10^{-6} < \alpha = 0{,}1$ sehingga dapat disimpulkan bahwa secara serentak minimal ada satu variabel bebas berpengaruh terhadap lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam angkatan 2011.

Setelah dilakukan pengujian parameter secara serentak, maka selanjutnya dilakukan pengujian signifikansi parameter secara individu dengan hipotesis sebagai berikut:

$H_0: \beta_j = 0$, dengan $j = 1, 2, 3, 4$

$H_1: \beta_j \neq 0$, dengan $j = 1, 2, 3, 4$

Berdasarkan Tabel 4, diperoleh *p-value* untuk variabel jenis kelamin dan IPK nilainya lebih kecil dibandingkan nilai $\alpha = 0{,}1$ sehingga dapat disimpulkan bahwa variabel jenis kelamin dan IPK signifikan berpengaruh terhadap lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam angkatan 2011 sehingga variabel jenis kelamin dan IPK dimasukkan ke dalam model. Kemudian dilakukan analisis kembali dengan mengeluarkan variabel yang tidak signifikan yaitu variabel daerah asal dan sumber pendanaan sebagai berikut:

Tabel 3. Hasil Estimasi Parameter Model Cox untuk Variabel Signifikan

Variabel	β	*p-value*	Keputusan
Jenis Kelamin	0,309	0,0734	H_0 ditolak
IPK	2,032	$8{,}8 \times 10^{-7}$	H_0 ditolak
Likelihood Ratio Test		$5{,}06 \times 10^{-7}$	H_0 ditolak

Berdasarkan Tabel 5, diperoleh *p-value* sebesar $5{,}06 \times 10^{-7} < \alpha = 0{,}1$ sehingga dapat disimpulkan bahwa secara serentak minimal ada satu variabel bebas berpengaruh terhadap lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam angkatan 2011.

Untuk pengujian secara individu, berdasarkan Tabel 5, diperoleh *p-value* untuk variabel jenis kelamin dan IPK nilainya lebih kecil dibandingkan nilai $\alpha = 0{,}1$, sehingga dapat disimpukan bahwa variabel jenis kelamin dan IPK signifikan berpengaruh terhadap waktu kelulusan mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam angkatan 2011. Sehingga diperoleh model regresi Cox *proportional hazard* sebagai berikut:

$$\hat{h}(t, \boldsymbol{x}) = \hat{h}_0(t)\exp(0{,}309x_1 + 2{,}032x_4) \quad (26)$$

Berdasarkan model yang diperoleh diketahui bahwa nilai *hazard ratio* untuk variabel jenis kelamin adalah $\widehat{\mathrm{HR}} = e^{(0{,}309)} = 1{,}362$. Artinya mahasiswa yang berjenis kelamin perempuan memiliki kesempatan 1,362 kali untuk lulus lebih cepat dibandingkan dengan mahasiswa yang berjenis kelamin laki-laki jika variabel lain dianggap konstan. Sedangkan nilai *hazard ratio* untuk penambahan IPK sebesar 0,1 adalah $\widehat{\mathrm{HR}} = e^{(0{,}1\times 2{,}032)} = 1{,}225$. Artinya setiap penambahan IPK sebesar 0,1, maka kesempatan mahasiswa untuk lulus lebih cepat akan meningkat sebesar 1,225 kali jka variabel lain dianggap konstan.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka diperoleh beberapa kesimpulan sebagai berikut:

Setelah dilakukan estimasi parameter dengan pendekatan Efron *partial likelihood* dan mengeluarkan variabel bebas yang tidak berpengaruh, maka diperoleh model regresi Cox *proportional hazard* untuk data lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 sampai dengan 28 Februari 2017 adalah sebagai berikut:

$$\hat{h}(t, \boldsymbol{x}) = h_0(t)\exp(0{,}309x_1 + 2{,}032x_4)$$

Berdasarkan hasil pengujian signifikansi parameter baik secara serentak maupun secara individu diperoleh bahwa variabel bebas yang berpengaruh signifikan terhadap lama masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman angkatan 2011 sampai dengan 28 Februari 2017 adalah variabel jenis kelamin dan IPK mahasiswa.

Berdasarkan model yang diperoleh diketahui bahwa nilai *hazard ratio* untuk variabel jenis kelamin adalah $\widehat{\mathrm{HR}} = e^{(0{,}309)} = 1{,}362$. Artinya mahasiswa yang berjenis kelamin perempuan memiliki kesempatan 1,362 kali untuk lulus lebih cepat dibandingkan dengan mahasiswa yang berjenis kelamin laki-laki jika variabel lain dianggap konstan. Sedangkan nilai *hazard ratio* untuk penambahan IPK sebesar 0,1 adalah $\widehat{\mathrm{HR}} = e^{(0{,}1\times 2{,}032)} = 1{,}225$. Artinya setiap penambahan IPK sebesar 0,1, maka kesempatan mahasiswa untuk lulus lebih cepat akan meningkat sebesar 1,225 kali jka variabel lain dianggap konstan.

Daftar Pustaka

Allison, P. (2010). *Survival Analysis Using SAS: A Practical Guide* (2nd Ed). Cary: SAS Institute Inc.

Collett, D. (2003). *Modelling Survival Data in Medical Research* (2nd Ed). London: Chapman & Hall/CRC.

Hosmer, D. W., dan Lemeshow, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Hoboken: John Wiley & Sons, Inc.

Klein, J. P., dan Moeschberger, M. L. (2003). *Survival Analysis: Techiques for Censored and Truncated Data* (2nd Ed). New York: Springer-Verlag.

Kleinbaum, D. G., dan Klein, M. (2005). *Survival Analysis: A Self-Learning Text* (2nd Ed).

New York: Springer Science+Business Media, Inc.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd Ed). Hoboken: John Wiley & Sons, Inc.

Lee, E. T., dan Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis* (3rd Ed). Hoboken: John Wiley & Sons, Inc.

Rahmadeni, Ranti, S. (2016). Perbandingan Model Cox Menggunakan Estimasi Efron *Partial Likelihood* dan Breslow *Partial Likelihood*. *Prosiding Seminar Nasional Teknologi Informasi, Komunikasi dan Industri Universitas Islam Negeri Sultan Syarif Kasim Riau*, Pekanbaru.

Vitriana, A. N., Kusumawati, R. (2016). Model Cox *Extended* dengan g(t)=1 untuk Mengatasi *Nonproportional Hazard* pada Kejadian Bersama. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika Universitas Negeri Yogyakarta*, Yogyakarta.