Analisis *Survival* pada Data Kejadian Bersama Menggunakan Metode *Exact Partial Likelihood* (Studi Kasus: Kecelakaan Lalu Lintas di Kota Samarinda Tahun 2016)

Survival Analysis Ties Event Data Using Exact Partial Likelihood Method (Case Study: Traffic Accident in Samarinda City 2016)

Rahmawati Isnaeni[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: rahmastat2013@gmail.com

Abstract

Survival data is the data of survival time until the appearance of certain events. In the survival analysis, ties are sometimes found, that is the situation where there are two or more individuals who experience the same event at the same time. There are several methods in estimating the parameters in the case of a ties, one of which is by applying the exact partial likelihood method. The exact method is the most accurate method, which can be applied in estimating Cox regression parameters from traffic accident data of Samarinda City in 2016. Traffic accidents are one of the most deadly events. The four main factors that cause traffic accidents are human factors, vehicles, roads, and weather or environmental factors. The variables used in this study are age, gender, role of victim, driver's license, vehicle type, time of incident, line of the road, and weather. The results of analysis with the help of Rstudio software showed that the factors whose affect the fatality rate of traffic accident victims of Samarinda City are age and gender. For the age variables concluded that each addition of one year of age of the accident victim, the risk of dying from a traffic accident will also increase 1,0258 times. As for the gender variables concluded that the victim of male sex has a risk of 0.4180 times greater to die due to traffic accidents compared with female victims.

Keywords: Cox proportional hazard, exact partial likelihood, survival, ties, traffic accident.

Pendahuluan

Analisis data uji hidup merupakan salah satu teknik statistika yang digunakan untuk menganalisis pengujian tentang waktu hidup satu unit atau komponen pada keadaan operasional tertentu. Data yang digunakan dalam analisis ini adalah data waktu hidup yang merupakan variabel random *non* negatif dan dapat berbentuk data lengkap atau data tidak lengkap. (Lee dan Wang, 2003). Analisis *survival* pada masa kini lebih banyak difokuskan pada fungsi *hazard* yaitu menganalisis peluang kejadian. Model Cox sering digunakan daripada metode lainnya karena dapat mengestimasi *hazard ratio* tanpa perlu diketahui fungsi *hazard* dasarnya. Waktu kejadian atau waktu *survival* dalam analisis *survival* terbagi menjadi 2 macam yaitu waktu kejadian tanpa *ties* dan waktu kejadian dengan *ties*. *Ties* atau kejadian bersama adalah keadaan dimana terdapat dua individu atau lebih yang mengalami kejadian pada waktu bersamaan. Peneliti sering menghindari adanya data yang memiliki kejadian bersama karena kejadian bersama mengakibatkan permasalahan dalam menyelesaikan persamaan *partial likelihood* (Hosmer, dkk, 2008).

Metode *exact* merupakan metode yang paling akurat dan dikenal paling dominan dalam memodelkan variabel daripada tiga metode lainnya seperti Breslow, *discrete* dan Efron yang pada tahapan iterasi dan eliminasi prediktor seringkali menghasilkan nilai residual yang besar dan besarnya pengaruh model yang lebih kecil (Therneau dan Grambsch, 2000). Penelitian ini akan menggunakan model Cox *Proportional Hazard* dengan pendekatan metode *exact* pada kasus kecelakaan lalu lintas. Kecelakaan lalu lintas merupakan salah satu peristiwa yang banyak memakan korban jiwa. Lembaga kesehatan dunia atau *World Health Organization* (WHO) baru-baru ini merilis *Global Status Report on Road Safety* yang menampilkan angka kecelakaan lalu lintas yang terjadi di 180 negara sepanjang tahun 2015. Indonesia menjadi negara ketiga di Asia di bawah Tiongkok dan India dengan total 38.279 total kematian akibat kecelakaan lalu lintas di tahun 2015 (WHO, 2016).

Di Kalimantan Timur khususnya kota Samarinda, tercatat dari Laka Lantas Polresta Samarinda jumlah kecelakaan yang terjadi pada tahun 2014 sebanyak 226 kasus dengan kerugian materil sebanyak Rp. 932.925.000,00. Sedangkan pada tahun 2015 jumlah kecelakaan yang terjadi adalah 133 kasus dengan kerugian materil sebanyak Rp. 326.850.000,00. Kendala yang dialami Satlantas pada umumnya dalam menekan angka kecelakaan lalu lintas adalah pada unsur masyarakat sebagai objek sekaligus subjek utama dari pengguna jalan (Dewi, 2016).

Berdasarkan hasil penelitian yang dilakukan oleh Susetyo (2014) yang membahas *survival* siswa putus sekolah data kejadian bersama dengan model Cox *proportional hazard* menggunakan metode *exact partial likelihood* memperlihatkan bahwa metode *exact partial likelihood* lebih baik dalam mengestimasi nilai parameter regresi dibanding metode Efron *partial likelihood*. Berdasarkan uraian tersebut penulis ingin membahas model Cox *propotional hazard* pada data kejadian bersama menggunakan metode *exact partial likelihood* untuk memodelkan penyebab kecelakaan yang terjadi di Kota Samarinda.

Analisis *Survival*

Analisis *survival* adalah salah satu cabang statistika yang mempelajari teknik analisis data *survival*. Data *survival* adalah data waktu bertahan sampai munculnya kejadian tertentu. Waktu *survival* adalah catatan waktu yang dicapai suatu objek sampai terjadinya peristiwa tertentu yang disebut sebagai *failure event* (Hosmer, dkk, 2008).

Kejadian Bersama

Dalam analisis *survival* terkadang ditemukan adanya kejadian bersama atau yang sering disebut *ties*. *Ties* adalah keadaan yang terdapat dua individu atau lebih yang mengalami kejadian pada waktu yang bersamaan. Jika pada suatu data terdapat *ties*, maka akan menimbulkan permasalahan dalam membentuk fungsi *partial likelihood*-nya yaitu saat menentukan anggota dari himpunan risikonya (Hosmer, dkk, 2008).

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

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

Misalkan T adalah variabel random bukan 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):

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

Fungsi *Survival*

Fungsi *survival* $S(t)$, mendefinisikan probabilitas dari suatu individu untuk bertahan setelah waktu yang ditetapkan, yang dinamakan t. Peluang suatu individu mampu hidup lebih lama daripada t adalah (Lee dan Wang, 2003):

$$\begin{aligned} S(t) &= P(T>t)=1-(P(T\le t) \\ &=1-\int_0^t f(x)\,dx=1-F(t) \end{aligned} \quad (3)$$

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, dkk, 2008):

$$\hat{S}(t)=\prod_{t_i<t}\frac{n_i-d_i}{n_i} \quad (4)$$

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

Fungsi *Hazard*

Jika T variabel acak non negatif pada interval $t=0$ menuju $t=\infty$ yang menunjukkan waktu hidup individu suatu populasi, maka risiko individu tersebut mengalami kegagalan pada interval t hingga $[t, t+\Delta t]$ dinyatakan dengan fungsi *hazard* $h(t)$ (Lee dan Wang, 2003).

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

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

$$H(t)=\int_0^t h(x)\,dx \quad (6)$$

Model Cox *Proportional Hazard*

Model Cox *proportional hazard* disebut dengan model Cox karena fungsi *hazard* dari individu yang berbeda adalah *proportional* atau rasio dari fungsi *hazard* dua individu yang berbeda adalah konstan (Lee dan Wang, 2003). Model Cox dapat dituliskan sebagai berikut:

$$h(t,\mathbf{x})=h_0(t)\exp\left(\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p\right) \quad (7)$$

dengan:

$h(t,x)$ = model Cox

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

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

$x_1,x_2,...,x_p$ = variabel bebas

Estimasi Parameter

Parameter β_j pada model Cox *proportional hazard* akan diestimasi dengan menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE). Fungsi *partial likelihood* memperhatikan peluang untuk subjek yang mengalami kejadian dan urutan kejadian. Pendugaan β_j dengan metode MPLE adalah nilai ketika fungsi *partial likelihood maximum*. Misal data untuk n individu yang terdiri dari r waktu kejadian yang tidak tersensor dan $n-r$ individu tersensor kanan, diurutkan menjadi $t_1 < t_2 < ... < t_i < ... < t_n$ dengan t_i merupakan urutan waktu kejadian ke-i. Diasumsikan hanya terdapat satu individu yang mengalami kematian pada tiap waktu kegagalan, jadi 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 (Hosmer, dkk, 2008).

$$P(A|B) = \frac{h_i(t_i)}{\sum_{l \in R(t_i)} h_l(t_i)} \quad (8)$$

Berdasarkan hasil peluang bersyarat pada Persamaan (8), maka diperoleh fungsi *partial likelihood* sebagai berikut:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p} \beta_j x_{ij})}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})} \quad (9)$$

Dari Persamaan (9) diperoleh fungsi *log partial likelihood* sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \ln \left[\prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p} \beta_j x_{ij})}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})} \right]$$
$$= \sum_{i=1}^{r} \left[\left(\sum_{j=1}^{p} \beta_j x_{ij} \right) - \left(\ln \sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \right) \right] \quad (10)$$

Estimasi Parameter dengan Pendekatan *Exact*

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

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp\left(\sum_{l \in D(t_i)} \sum_{j=1}^{p} x_{lj}(t_i) \beta_j \right)}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^{p} x_{lj}(t_i) \beta_j \right)} \quad (11)$$

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

$l = \{l_1, l_2, ..., l_{d_i}\}$ = elemen dari $R(t_i)$

d_i = banyaknya *ties* yang teramati pada saat t_i

Dari Persamaan (11) diperoleh fungsi *log partial likelihood* sebagai berikut:

$$\ln L(\boldsymbol{\beta}) = \ln \prod_{i=1}^{r} \frac{\exp\left(\sum_{l \in D(t_i)} \sum_{j=1}^{p} x_{lj}(t_i) \beta_j \right)}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^{p} x_{lj}(t_i) \beta_j \right)}$$
$$= \sum_{i=1}^{r} \left[\begin{array}{l} \sum_{l \in D(t_i)} \sum_{j=1}^{p} x_{lj}(t_i) \beta_j \\ -\ln\left(\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^{p} x_{lj}(t_i) \beta_j \right) \right) \end{array} \right] \quad (12)$$

Pengujian Asumsi *Proportional Hazard*

Pengujian asumsi *proportional hazard* secara statistik adalah dengan menggunakan uji *goodness of fit*. Langkah-langkah pengujian asumsi *proportional hazard* dengan uji *goodness of fit* adalah sebagai berikut (Lee dan Wang, 2003):

a. Menggunakan model Cox *proportional hazard* untuk mendapatkan residual Schoenfeld untuk setiap variabel bebas. Residual Schoenfeld untuk individu ke-*i* pada variabel bebas ke-*j* adalah sebagai berikut:

$$R_{ji} = \delta_i \left(x_{ji} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp\left(\hat{\beta}' x_l \right)}{\sum_{l \in R(t_i)} \exp\left(\hat{\beta}' x_l \right)} \right) \quad (13)$$

dengan:

R_{ji} = taksiran residual Schoenfeld dari variabel *j* untuk individu ke-*i*

$R(t_i)$ = himpunan individu yang berisiko pada waktu t_i

δ_i = indikator sensoring untuk individu ke-*i*

x_{ji} = nilai dari variabel *j* untuk individu ke-*i* dengan $j = 1, 2, ..., p$ dan $i = 1, 2, ..., n$

b. Mengurutkan waktu individu mulai dari individu yang mengalami *event* pertama kali.

c. Menguji korelasi antara variabel residual Schoenfeld dan urutan waktu *survival* menggunakan korelasi Pearson sebagai berikut:

$$r = \frac{\sum_{i=1}^{n} \left(R_{ji} - \overline{R_{ji}} \right) \left(y_i - \overline{y_i} \right)}{\sqrt{\sum_{i=1}^{n} \left(R_{ji} - \overline{R_{ji}} \right)^2 \sum_{i=1}^{n} \left(y_i - \overline{y_i} \right)^2}} \quad (14)$$

dengan:

r = nilai korelasi antara residual Schoenfeld dengan urutan waktu

$\overline{R_{ji}}$ = rata-rata taksiran residual Schoenfeld dari variabel *j* untuk individu ke-i

y_i = urutan waktu *survival*, $i = 1, 2, ..., n$

$\overline{y_i}$ = rata-rata urutan waktu *survival*

Dengan pengujian hipotesis sebagai berikut:

Hipotesis

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

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

Statistik uji

$$z_{hitung} = \frac{r}{\frac{1}{\sqrt{n-1}}} \quad (15)$$

dimana n adalah banyak individu yang diamati. Tolak H_0 jika $z_{hitung} > z_{(\alpha/2)}$ atau $p-value < \alpha$ yang berarti terdapat korelasi antara residual Schoenfeld dengan *rank* waktu *survival* sehingga asumsi *proportional hazard* tidak terpenuhi.

Pengujian Parameter

1. Uji *Likelihood Ratio*

Untuk mengetahui pengaruh seluruh variabel bebas dalam model secara bersama sama maka diperlukan suatu uji simultan. Uji simultan dilakukan secara uji *goodness of fit* untuk mengidentifikasi peubah yang berpotensi masuk dalam komponen linier model (Collet, 2003). Hipotesis dari pengujian ini adalah:

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

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

Statistik uji

$$G = -2\left[\ln L(0) - \ln L(\hat{\beta})\right] \quad (16)$$

di mana:

$\ln L(0)$ = *log likelihood* dari model tanpa variabel bebas.

$\ln L(\hat{\beta})$ = *log likelihood* dari model yang terdiri dari p variabel bebas.

Uji ini mengikuti sebaran *chi-square* dengan derajat bebas p, dimana H_0 ditolak jika nilai $G \geq x^2_{(\alpha,p)}$ atau $p-value < \alpha$ sehingga dapat disimpulkan secara simultan semua variabel bebas berpengaruh terhadap variabel terikat.

2. Uji Wald

Uji Wald digunakan untuk menguji pengaruh estimasi parameter masing-masing variabel bebas terhadap variabel terikat (Collet, 2003). Hipotesis dalam pengujian ini adalah:

$H_0 : \beta_k = 0$, dengan $j = 1, 2, ..., p$

$H_0 : \beta_k \neq 0$, dengan $j = 1, 2, ..., p$

Statistik uji

$$W = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}\right)^2 \quad (17)$$

di mana:

W = statistik uji Wald

$\hat{\beta}_k$ = koefisien variabel bebas ke-k

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

Uji ini mengikuti sebaran *Chi-square* dengan derajat bebas $df = p$, dimana H_0 ditolak jika nilai $W > x^2_{(\alpha,1)}$ atau $p-value < \alpha$ sehingga dapat disimpulkan secara parsial variabel bebas berpengaruh terhadap variabel terikat.

Hasil dan Pembahasan

Estimasi Parameter dengan *Exact*

Langkah pertama yang akan dilakukan adalah membentuk fungsi *partial likelihood* untuk metode *exact* sebagai berikut:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{107} \frac{\exp\left(\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)} \quad (18)$$

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

$$\ln L(\boldsymbol{\beta}) = \ln \prod_{i=1}^{107} \frac{\exp\left(\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)}$$

$$= \sum_{i=1}^{107}\left[\begin{array}{l}\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i)\beta_j \\ -\ln\left(\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\right)\end{array}\right] \quad (19)$$

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

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial \sum_{i=1}^{107}\left[\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i)\beta_j - \ln \sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\right]}{\partial \beta_j}$$

$$= \sum_{i=1}^{107}\left[\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i) - \frac{\sum_{l\in R(t_i)}\exp\left(\sum_{i=1}^{8} x_{lj}(t_i)\beta_j\right)\left(\sum_{j=1}^{8} x_{lj}(t_i)\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{i=1}^{8} x_{lj}(t_i)\beta_j\right)}\right] \quad (20)$$

Turunan parsial kedua dari $\ln L(\boldsymbol{\beta})$ metode *exact* adalah sebagai berikut:

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial^2 \ln \beta_j} = \frac{\partial}{\partial \beta_j}\left[\sum_{i=1}^{107}\left[\sum_{l\in D(t_i)}\sum_{j=1}^{8} x_{lj}(t_i) - \frac{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\left(\sum_{j=1}^{8} x_{lj}(t_i)\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)}\right]\right]$$

$$= \sum_{i=1}^{107}\left[\begin{array}{l}\frac{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\left(\sum_{j=1}^{8} x_{lj}(t_i)\right)\right]^2}{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\right]^2} - \\ \frac{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\left(\sum_{j=1}^{8} x_{lj}(t_i)\right)^2\right]}{\left[\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{8} x_{lj}(t_i)\beta_j\right)\right]}\end{array}\right] \quad (21)$$

Berdasarkan hasil turunan pertama yang diperoleh dari metode *exact*, dapat dilihat bahwa terdapat fungsi yang tidak dapat diselesaikan secara analitis untuk mendapatkan hasil estimasi masing-masing parameter. Pada kasus seperti ini, maka suatu 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 dengan bantuan komputasi. Iterasi Newton-Raphson untuk mengestimasi parameter model Cox dilakukan dengan bantuan *software* R.

Gambaran Umum Data

Data yang digunakan dalam penelitian ini adalah data korban kecelakaan lalu lintas Kota Samarinda tahun 2016, dengan variabel terikatnya yaitu waktu *survival* yang dinyatakan dalam satuan hari dan berada dalam batas periode penelitian dengan ketentuan waktu *survival* dikategorikan teramati atau tidak tersensor $\delta_i = 1$ jika korban kecelakaan lalu lintas ke-*i* meninggal dunia, sedangkan waktu survival dikategorikan tidak teramati atau tersensor $\delta_i = 0$ jika korban kecelakaan lalu lintas tetap hidup setelah kecelakaan.

Adapun variabel bebas yang digunakan adalah sebagai berikut:

a. Usia Korban (x_1)

Usia korban merupakan usia korban saat terjadi kecelakaan dan berskala data kontinu.

b. Jenis Kelamin (x_2)

Jenis kelamin korban dikategorikan menjadi:

$$x_2 = \begin{cases} 0, & \text{jika korban adalah perempuan} \\ 1, & \text{jika korban adalah laki-laki} \end{cases}$$

c. Peran Korban (x_3)

Peran korban dikategorikan menjadi:

$$x_3 = \begin{cases} 0, & \text{jika peran korban adalah selain pengemudi} \\ 1, & \text{jika peran korban adalah pengemudi} \end{cases}$$

d. Kepemilikan SIM (x_4)

Kepemilikan SIM dikategorikan menjadi:

$$x_4 = \begin{cases} 0, & \text{jika korban tidak memiliki SIM} \\ & \quad \text{atau tidak memerlukan SIM} \\ 1, & \text{jika korban memiliki SIM} \end{cases}$$

e. Jenis Kendaraan (x_5)

Jenis kendaraan dalam penelitian ini hanya terfokus pada dua jenis kendaraan bermotor, dan lebih menitikberatkan kepada jenis kendaraan roda dua yang dianggap lebih berisiko menyebabkan penggunanya meninggal dunia ketika terjadi kecelakaan dibandingkan kendaraan roda empat. Jenis kendaraan dikategorikan sebagai berikut:

$$x_5 = \begin{cases} 0, & \text{jika selain kendaraan roda dua} \\ 1, & \text{jika kendaraan roda dua} \end{cases}$$

f. Waktu Kejadian (x_6)

Variabel waktu kejadian yang dimaksud ialah kapan terjadinya kecelakaan. Waktu kejadian digolongkan menjadi 2, yaitu:

$$x_6 = \begin{cases} 0, & \text{jika terjadi pada pukul} \\ & \quad 06.00 \le x_6 < 18.00 \text{ WITA} \\ 1, & \text{jika terjadi pada pukul} \\ & \quad 18.00 \le x_6 < 06.00 \text{ WITA} \end{cases}$$

g. Pola Arus (x_7)

Pola arus digolongkan menjadi 2, yaitu:

$$x_7 = \begin{cases} 0, & \text{jika terjadi dalam pola arus 1 arah} \\ 1, & \text{jika terjadi dalam pola arus 2 arah} \end{cases}$$

h. Cuaca (x_8)

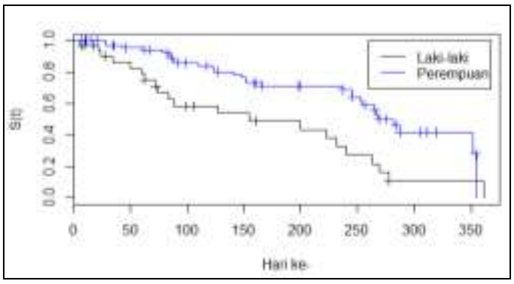
Cuaca digolongkan menjadi 2, yaitu:

$$x_8 = \begin{cases} 0, & \text{jika terjadi ketika cuaca cerah} \\ 1, & \text{jika terjadi ketika cuaca hujan} \end{cases}$$

Grafik *Survival*

Grafik *survival* Kaplan Meier digunakan untuk mengetahui deskriptif karakteristik grafik *survival* korban kecelakaan lalu lintas berdasarkan faktor-faktor kategorik yang diduga mempengaruhi tingkat kecelakaan lalu lintas yaitu jenis kelamin, peran korban, kepemilikan SIM, jenis kendaraan, waktu kejadian, pola arus, dan cuaca. Estimasi fungsi *survival* dengan metode Kaplan-Meier dilakukan dengan bantuan *software* R berdasarkan Persamaan (4).

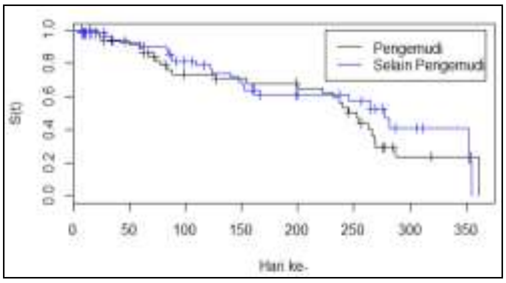
Karakteristik berdasarkan variabel jenis kelamin sebagai berikut:



Gambar 1. Grafik *survival* Kaplan-Meier jenis kelamin

Berdasarkan Gambar 1 dapat dilihat grafik *survival* Kaplan Meier menampilkan pada hari ke-22 hingga hari ke-355 terlihat bahwa grafik untuk jenis kelamin laki-laki berada di bawah grafik untuk jenis kelamin perempuan, artinya korban yang berjenis kelamin laki-laki memiliki peluang meninggal dunia lebih tinggi dibanding korban yang berjenis kelamin perempuan ketika mengalami kecelakaan.

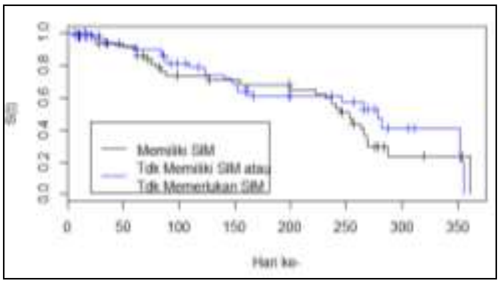
Karakteristik berdasarkan variabel peran korban sebagai berikut:



Gambar 2. Grafik *survival* Kaplan Meier peran korban

Berdasarkan Gambar 2 dapat dilihat grafik *survival* Kaplan Meier menampilkan bahwa pada hari ke-143 hingga hari ke-237 terlihat bahwa grafik korban yang merupakan pengemudi berada di atas grafik korban yang berperan sebagai selain pengemudi, artinya korban yang berperan sebagai pengemudi memiliki peluang meninggal dunia lebih tinggi dibanding korban yang berperan sebagai selain pengemudi ketika mengalami kecelakaan.

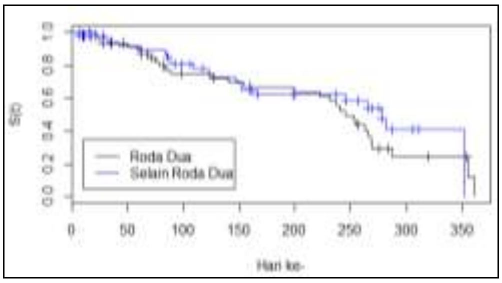
Karakteristik berdasarkan variabel kepemilikan SIM sebagai berikut:



Gambar 3. Grafik *survival* Kaplan Meier kepemilikan SIM

Berdasarkan Gambar 3 dapat dilihat grafik *survival* Kaplan Meier menampilkan bahwa pada hari ke-122 hingga hatri ke-243 terlihat bahwa grafik korban yang memiliki SIM berada di atas grafik korban yang tidak memiliki SIM atau tidak memerlukan SIM, artinya korban yang memiliki SIM mempunyai peluang meninggal dunia lebih tinggi dibanding korban yang tidak memiliki SIM atau tidak memerlukan SIM ketika mengalami kecelakaan. Sedangkan pada hari ke-62 hingga hari ke-87 dan hari ke-244 hingga heri ke-352, terlihat bahwa grafik korban yang memiliki SIM berada di bawah grafik korban yang tidak memiliki SIM atau tidak memerlukan SIM, artinya korban yang memiliki SIM mempunyai peluang meninggal dunia lebih rendah dibanding korban yang tidak memiliki SIM atau tidak memerlukan SIM ketika mengalami kecelakaan.

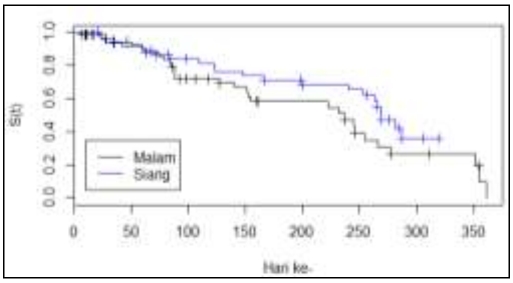
Karakteristik berdasarkan variabel jenis kendaraan sebagai berikut:



Gambar 4. Grafik *survival* Kaplan Meier jenis kendaraan

Berdasarkan Gambar 4 dapat dilihat grafik *survival* Kaplan Meier menampilkan bahwa pada hari ke-150 hingga ke-233, grafik untuk korban yang menggunakan kendaraan roda dua berada di atas grafik untuk korban yang menggunakan kendaaraan selain roda dua, artinya korban kecelakaan yang menggunakan kendaraan roda dua mempunyai peluang meninggal dunia lebih tinggi dibanding korban kecelakaan yang menggunakan kendaraan selain roda dua. Sedangkan untuk hari ke-54 hingga hari ke-137, dan pada hari ke-234 hingga hari ke-350, terlihat bahwa grafik korban kecelakaan yang menggunakan kendaraan roda dua berada di bawah grafik korban kecelakaan yang menggunakan kendaraan selain roda dua, artinya korban kecelakaan yang menggunakan kendaraan roda dua mempunyai peluang meninggal dunia lebih rendah dibanding korban kecelakaan yang menggunakan kendaraan selain roda dua ketika mengalami kecelakaan.

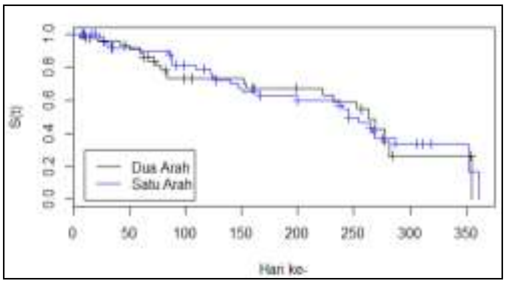
Karakteristik berdasarkan variabel waktu kejadian sebagai berikut:



Gambar 5. Grafik *survival* Kaplan Meier waktu kejadian

Berdasarkan Gambar 5 dapat dilihat pada hari ke-73 hingga hari ke-361 terlihat bahwa grafik untuk korban kecelakaan pada malam hari berada di bawah grafik korban kecelakaan pada siang hari, artinya korban kecelakaan pada malam hari mempunyai peluang meninggal dunia lebih rendah dibanding korban kecelakaan pada siang hari ketika mengalami kecelakaan.

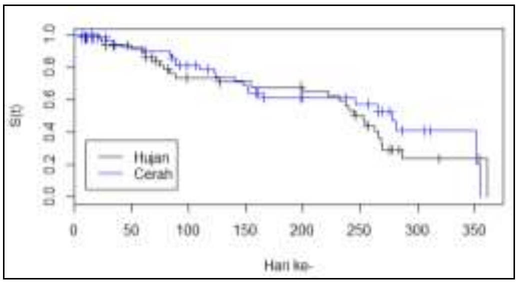
Karakteristik berdasarkan variabel pola arus sebagai berikut:



Gambar 6. Grafik *survival* Kaplan Meier pola arus

Berdasarkan Gambar 3 dapat dilihat pada hari ke-143 hingga ke-270, terlihat bahwa grafik untuk korban kecelakaan pada pola arus dua arah berada diatas grafik untuk korban kecelakaan pada pola arus satu arah saling, artinya korban kecelakaan pada pola arus dua arah memiliki peluang yang lebih tinggi untuk meninggal dunia ketika mengalami kecelakaan daripada korban kecelakaan pada pola arus satu arah. Sedangkan untuk hari ke-57 hingga hari ke-142 dan pada hari ke-271 hingga ke-350, terlihat bahwa grafik korban kecelakaan pada pola arus dua arah berada di bawah grafik korban kecelakaan pada pola arus satu arah, artinya korban kecelakaan pada pola arus dua arah mempunyai peluang meninggal dunia lebih rendah dibanding korban kecelakaan pada pola arus satu arah ketika mengalami kecelakaan.

Karakteristik berdasarkan variabel cuaca sebagai berikut:



Gambar 7. Grafik *survival* Kaplan Meier cuaca

Berdasarkan Gambar 7 dapat dilihat pada hari ke-32 hingga hari ke-73, dan pada hari ke-240 hingga hari ke-361, grafik korban kecelakaan pada saat cuaca hujan berada di atas grafik korban kecelakaan pada saat cuaca cerah, artinya korban kecelakaan pada saat cuaca hujan mempunyai peluang meninggal dunia lebih tinggi dibanding korban kecelakaan pada saat cuaca cerah. Sedangkan pada hari ke-33 hingga hari ke-200, grafik korban kecelakaan pada saat cuaca hujan berada di bawah grafik korban kecelakaan pada saat cuaca cerah, artinya korban kecelakaan pada saat cuaca hujan mempunyai peluang meninggal dunia lebih rendah dibanding korban kecelakaan pada saat cuaca cerah ketika mengalami kecelakaan.

Pengujian Asumsi *Proportional Hazard*

Pengujian asumsi *proportional hazard* dilakukan dengan uji *goodness of fit* menggunakan residual Schoenfeld pada persamaan (13). Adapun hasil uji *goodness of fit* dengan bantuan *software* R sebagai berikut:

Tabel 1. Hasil Uji *Goodness of Fit*

Variabel	*p-value*	Keputusan
Jenis Kelamin	0,117	H_0 gagal ditolak
Peran Korban	0,909	H_0 gagal ditolak
Kepemilikan SIM	0,760	H_0 gagal ditolak
Jenis Kendaraan	0,425	H_0 gagal ditolak
Waktu Kejadian	0,907	H_0 gagal ditolak
Pola Arus	0,814	H_0 gagal ditolak
Cuaca	0,139	H_0 gagal ditolak

Hasil uji *goodness of fit* sesuai pada Tabel 1 dengan menggunakan $\alpha = 0,05$, diperoleh bahwa semua *rank* waktu *survival* variabel bebas kategorik yaitu variabel jenis kelamin, peran korban, kepemilikan SIM, jenis kendaraan, waktu kejadian, pola arus, dan cuaca tidak memiliki korelasi terhadap masing-masing residual Schoenfeldnya, sehingga asumsi *proportional hazard* semua variabel bebas kategorik terpenuhi. Untuk itu selanjutnya, dapat digunakan model regresi Cox *proportional hazard*.

Estimasi Parameter Regresi

Setelah melakukan pengujian asumsi *proportional hazard*, tahap selanjutnya adalah pembentukan model regresi Cox *proportional hazard* untuk mengetahui variabel bebas yang berpengaruh terhadap waktu *survival* korban kecelakaan lalu lintas Kota Samarinda tahun 2016. Hasil estimasi parameter model regresi Cox *proportional hazard* dengan pendekatan *exact* dengan bantuan *software* R sebagai berikut:

Tabel 2. Hasil Estimasi dan Pengujian Model Sementara

Variabel	β	*Hazard Ratio*	p-*Value*	Keputusan
Usia	0,0278	1,0281	0,0019	
Jenis Kelamin	-0,8425	0,4306	0,0264	H_0 gagal ditolak
Peran Korban	0,2986	1,3480	0,6467	H_0 gagal ditolak
Kepemilikan SIM	-0,3289	0,7197	0,4333	H_0 gagal ditolak
Jenis Kendaraan	-0,1980	0,8203	0,7173	H_0 gagal ditolak
Waktu Kejadian	-0,1748	0,8396	0,5741	H_0 gagal ditolak
Pola Arus	0,2655	1,3041	0,4178	H_0 gagal ditolak
Cuaca	0,4739	1,6064	0,2083	H_0 gagal ditolak
Likelihood ratio			0,0056	Tolak H_0

Dengan menggunakan $\alpha = 0,05$, nilai *likelihood ratio test* sebesar 0,0056 berarti bahwa secara simultan paling sedikit ada satu variabel bebas berpengaruh terhadap model Cox sementara. Sedangkan secara parsial variabel usia dan jenis kelamin masing-masing berpengaruh signifikan terhadap model dan variabel peran korban, kepemilikan SIM, jenis kendaraan, waktu kejadian, pola arus, dan cuaca masing-masing tidak berpengaruh signifikan terhadap model Cox sementara.

Karena terdapat beberapa variabel bebas yang tidak berpengaruh signifikan maka dilakukan pengujian ulang tanpa mengikutsertakan semua variabel bebas yang tidak berpengaruh. Hasil analisis estimasi β adalah sebagai berikut:

Tabel 3 Hasil Estimasi dan Pengujian Parameter Model Cox

Variabel	β	***Hazard Ratio***	**p-*Value***	**Keputusan**
Usia	0,0254	1,0258	0,0018	
Jenis Kelamin	-0,8722	0,4180	0,0034	H_0 gagal ditolak
Peran Korban	0,2986	1,3480	0,6467	H_0 gagal ditolak
Likelihood ratio			0,00000102	Tolak H_0

Dengan menggunakan $\alpha = 0,05$, nilai *likelihood ratio test* sebesar 0,00000102 berarti bahwa secara simultan paling sedikit ada satu variabel bebas berpengaruh terhadap model Cox. Sedangkan secara parsial variabel usia dan jenis kelamin masing-masing berpengaruh signifikan terhadap model Cox. Dari hasil analisis estimasi parameter dan pengujian parameter didapat model Cox sebagai berikut:

$$\hat{h}(t,x) = \hat{h}_0(t)\exp(0,0254x_1 - 0,8722x_2)$$

dimana x_1 untuk variabel usia korban dan x_2 untuk variabel jenis kelamin. Dengan menggunakan persamaan (2.17), dari model dapat dilihat bahwa nilai rasio *hazard* untuk variabel usia adalah 1,0258. Artinya setiap penambahan satu tahun usia korban kecelakaan, risiko untuk meninggal dunia akibat kecelakaan lalu lintas juga akan meningkat 1,0258 kali ketika variabel lain dianggap konstan. Dengan mengunakan persamaan (2.16), diperoleh nilai rasio *hazard* untuk variabel jenis kelamin adalah 0,4180. Artinya korban yang berjenis kelamin laki-laki memiliki risiko 0,4180 kali untuk meninggal dunia akibat kecelakaan lalu lintas dibandingkan dengan korban yang berjenis kelamin perempuan ketika variabel lain dianggap konstan.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka diperoleh kesimpulan model regresi Cox terbaik untuk data *survival* dengan kejadian bersama pada korban kecelakaan lalu lintas di Kota Samarinda tahun 2016 menggunakan metode *exact* adalah sebaagai berikut:

$$\hat{h}(t,x) = \hat{h}_0(t)\exp(0,0254x_1 - 0,8722x_2)$$

dimana x_1 untuk variabel usia korban dan x_2 untuk variabel jenis kelamin. Dari model dapat dilihat bahwa nilai rasio *hazard* untuk variabel usia adalah 1,0258. Artinya setiap penambahan satu tahun usia korban kecelakaan, risiko untuk meninggal dunia akibat kecelakaan lalu lintas juga akan meningkat 1,0258 kali ketika variabel lain dianggap konstan. Nilai rasio *hazard* untuk variabel jenis kelamin adalah 0,4180. Artinya korban yang berjenis kelamin laki-laki memiliki risiko 0,4180 kali untuk meninggal dunia akibat kecelakaan lalu lintas dibandingkan dengan korban yang berjenis kelamin perempuan ketika variabel lain dianggap konstan.

Daftar Pustaka

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

Dewi, R. (2016). Hubungan Antara Tayangan 86 di Net dengan Tingkat Pengetahuan Peraturan Lalu Lintas (Studi Kasus pada Masyarakat Bubuhan Samarinda). *eJournal Ilmu Komunikasi Fakultas Ilmu Sosial dan Politik Universitas Mulawarman, 4(4).*

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

Susetyo, A. B. (2014). Analisis *Survival* Data Kejadian *Ties* dengan *Exact Partial Likelihood* pada *Cox Regression*; Studi Kasus Data Siswa Putus Sekolah Tingkat Menengah Pertama. *Prosiding Seminar Nasional Matematika,Universitas Jember*.

Therneau, T. M. dan Grambsch, P. M. (2000). *Modeling Survival Data: Extending The Cox Model*. New York: Springer-Verlag, Inc.

World Health Organization. (2016). *Global Status Report on Road Safety 2015*. Geneva: WHO Press.