Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 8 *ISSN : 2085-9902*
Pekanbaru, 9 November 2016

Perbandingan Model Regresi *Cox* Menggunakan Estimasi Paramater *Efron Partial Likelihood* dan *Breslow Partial Likelihood*

Rahmadeni[1], Syofia Ranti[2]
[1,2]Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
email: [1]r4dieni@gmail.com, [2]syofia.ranti94@gmail.com

Abstrak

Model Regresi cox proportional hazard pada kejadian bersama adalah modifikasi dari model regresi cox ketika terdapat dua atau lebih individu yang mengalami kejadian bersama. Model regresi cox proportionalhazard pada kasus ketahanan hidup pasien diabetes ini diestimasi dengan menggunakan dua pendekatan yaitu breslow partial likelihood dan efron partial likelihood, dimana kedua pendekatan ini merupakan modifikasi dari estimasi parameter menggunakan maksimum partial likelihood.Hasil estimasi parameter $\beta_1, \beta_2, \ldots \beta_p$ *diperoleh dengan memaksimalkan fungsi partial likelihood dengan menggunakan iterasi pada metode newton raphson. Model cox yang diperoleh yaitu model cox yang terdiri dari variabel CHD dengan menggunakan estimasi parameter efron partial likelihood dengan seleksi backward dilihat dari nilai AIC terkecil.*

Kata kunci : *metode Breslow dan Efron, model cox proportional hazard, seleksi backward.*

Abstract

Cox proportional hazard regression model at together event is a modification of cox regression model when be able two or more individual that have together event. The regression model of Cox proportional hazard at survival patient medical casewill be estimateby using two approachment, are breslow partial likelihood and efron partial likelihood, where both ofthis approachment are modification of parameter estimationby using maxsimum partial likelihood. The Result of parameter estimation $\beta_1, \beta_2, \ldots \beta_p$*obtainedby maximum partial likelihood function using newton raphsoniterationmethod. The obtained Cox model is cox model that comprised CHD variable with used efron partial likelihood parameter estimation with backward selection that seen in by the minimum AIC value.*

Key word : *backward selection**,** Breslow dan Efron method, cox proportional hazard model*

1. Pendahuluan

Analisis *survival* adalah sekumpulan prosedur statistika yang digunakan untuk menganalisa data dimana variabel yang diperhatikan adalah waktu sampai terjadinya suatu *event* [2].Dalam analisis *survival* terdapat variabel waktu sebagai waktu uji hidup *(survival time)*, waktu dapat dinyatakan dalam tahun, bulan, minggu dan hari dari awal mula dilakukan penelitian atau pengamatan pada seorang individu sampai terjadinya suatu peristiwa pada individu tersebut.Peristiwa yang dimaksud adalah perkembangan suatu penyakit, reaksi terhadap suatu percobaan, kambuhnya suatu penyakit dan kematian.Analisis ini dapat dilakukan dengan menggunakan regresi *cox proportional hazard.*
Regresi *cox proportional hazard* digunakan untuk menetukan besarnya hubungan antara variabel independen dengan variabel dependen.data yang digunakan adalah data diabetes yang memiliki waktu *ties* atau kejadian bersama, dan dengan pendekatan *Efron Partial Likelihood* dan *Breslow Partial Likelihood.*

Proses Identifikasi data dalam analsis survival sangatlah penting, dimana waktu survival serta variabel bebas dan terikat yang akan digunakan dalam penelitian ini ditentukan diawal proses analisis. Data yang digunakan didalam model regresi *Cox* ini yaitu contoh data yang diambildaribuku lee &wang, yaitu data pasien diabetes yang telahditelitioleh lee et padatahun 1988.

2. Analisis Data

1). Sebelum melakukan analisis lebih lanjut, maka uji asumsi *proportional hazard* perlu dilakukan untuk melihat apakah variable independen memenuhi *proportional hazard* atau tidak.
2). Melakukanestimasi parameter dengan menggunakan*Breslow* estimasi dan *Efron* estimasi.
3). Membentuk model Regresi *Cox* dengan memasukan nilai parameter yang diperoleh dari proses estimasi.
4). Memilih model terbaik pada Regresi *Cox* dengan menggunakan seleksi *Backward* dengan melihat nilai AIC yang terkecil.
5). Membandingkan kedua model regresi yang diperoleh dari estimasi parameter menggunakan *breslow* dan *efron* dengan melihat nilai AIC terkecil yang diperoleh dari masing-masing estimasi.
6). Menguji parameter dari model regresi *Cox* terbaik yang telah diperoleh dengan menggunakan Uji *Wald*dan Uji *Partial Like lihood* Rasio.
7). Menyusun model regresi *Cox* dengan menggunakan parameter terbaik yang telah diperoleh.
8). Menginterpretasikan model regresi *Cox* yang telah diperoleh dengan beberapa langkah diatas.

2.2 Analisis Survival

Analisis survival digunakan untuk menganalisis data waktu antar kejadian (*time to event* data) atau menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* sampai terjadinya suatu peristiwa khusus.Peristiwa khusus tersebut dapat berupa kejadian positif seperti kelahiran, kelulusan sekolah, kesembuhan dari suatu penyakit [4].

2.3 Fungsi *Survival*

Misalkan T adalah waktu bertahan hidup sampai munculnya suatu kejadian. Fungsi *survivalS(t)*, mendefinisikan probabilitas suatu individu untuk bertahan setelah waktu yang ditetapkan, namakan *t* [6].

$$S(t) = P(suatu\ individu\ untuk\ bertahan\ sampai\ waktu\ t)$$

$$S(t)P(T > t) \quad (1)$$

Fungsi survival dapat juga ditulis sebagai berikut :

$$S(t) = 1 - F(t), dimana\ F(t) = P(T \leq t) \quad (2)$$

2.4 Fungsi Kepadatan Peluang

Fungsi kepadatan peluang adalah peluang kegagalan suatu individu dalam interval waktu kecil per satuan waktu. Fungsi kepadatan peluang dinotasikan dengan $f(t)$ dan dirumuskan dengan:

$$f(x) = \lim_{\Delta t \to 0} \frac{P[objek\ gagal\ pada\ interval(t, t + \Delta t)]}{\Delta t}$$
$$= \lim_{\Delta t \to 0} [\frac{P(t<T<(t+\Delta t))}{\Delta t}] \quad (3)$$

2.5 Fungsi *Hazard*

Fungsi *hazard* didefinisikan sebagai *rate* suatu individu untuk mengalami *even*t dalam interval waktu dari t sampai $t + \Delta t$ jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t.

Hazard rate dapat didefinisikan sebagai berikut:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | \mathrm{T} \geq \mathrm{t})}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{P\big(t \leq T < (t + \Delta t)\big) \cap (\mathrm{T} \geq \mathrm{t})}{\Delta t. P(\mathrm{T} \geq \mathrm{t})}$$
$$= \lim_{\Delta t \to 0} \frac{P\big(t \leq T < (t + \Delta t)\big)}{\Delta t. P(\mathrm{T} \geq \mathrm{t})}$$

$$= \lim_{\Delta t \to 0} \frac{P\big(t \le T < (t + \Delta t)\big)}{\Delta t . S(t)}$$
$$= \frac{1}{S(t)} . \lim_{\Delta t \to 0} \frac{P\big(t \le T < (t + \Delta t)\big)}{\Delta t}$$
$$= \frac{1}{S(t)} . f(t)$$
$$\mathrm{h(t)} = \frac{\mathrm{f(t)}}{\mathrm{S(t)}} \qquad (4)$$

2.7 Model *Cox Proportional Hazard*

Model *Cox Proportional Hazard* disebut dengan model *cox* karena asumsi *proportionalhazardnya* yaitu fungsi *hazard* dari individu yang berbeda adalah *proportional* atau rasio dari fungsi *hazard* dua individu yang berbeda adalah konstan [6].
Model *hazard* proporsional umum adalah sebagai berikut:

$$h(t;x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p) \qquad (5)$$

dengan $h_0(t)$: Fungsi dasar *hazard* $\beta_1, \beta_2, ..., \beta_p$: Parameter regresi
$x_1, x_2, ..., x_p$: Nilai dari variabel bebas $X_1, X_2, ..., X_P$

Model hazard proporsional dapat pula dinyatakan dalam bentuk persamaan log yaitu:

$$\log h(t;x) = \log h_0(t)(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)$$
$$\log \frac{h(t,x)}{h_0(t)} = (\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p) \qquad (6)$$

2.8 Pengujian Asumsi *Proportional Hazard*

Ada 3 pilihan untuk mengatasi *Coxnonproportional hazard* yaitu mengeluarkan variabel bebas yang tidak memenuhi asumsi dari model, menggunakan model *Cox Stratifikasi* dan dengan perluasan model *Cox* [1].

Residual Schoenfeld dapat digunakan untuk mengecek asumsi *propotional hazard*. Nilai harapan dari *Scaled Shoenfeld Residual* ke-i, untuk variabelpenjelas ke-$j(X_j)$ dalam model r^*_{pji} diberikan oleh $E(r^*_{pji}) \approx \beta_j(t_i) - \hat{\beta}_j$ dimana $\beta_j(t_i)$ adalah nilai parameter pada saat waktu kejadian t_i. $\hat{\beta}_j$ merupakan nilai estimasi dari β_j pada model regresi cox. Plot dari nilai $r*_{pij} + \hat{\beta}_j$ berlawanan terhadap waktu kejadian memberikan informasi tentang bentuk dari koefisien dependen waktu dari X_j yaitu $\beta_j(t)$.

2.9 Estimasi Parameter Tanpa Kejadian Bersama

Fungsi *Likelihood* dari model regresi *cox proportional hazard* adalah sebagai berikut:

$$L(\beta) = \prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p} \beta_j x_{ij})}{\sum_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})} \qquad (7)$$

Dari fungsi *partial likelihood* diatas diperoleh fungsi *log-likelihood* sebagai berikut:

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

$$= \sum_{i=1}^{r}\left[(\sum_{j=1}^{p} \beta_j x_{ij}) - \log(\sum\nolimits_{l \varepsilon R(t_j)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})) \right] \quad (8)$$

Pendugaan parameter β_j dapat diperoleh dengan cara memaksimumkan turunan pertama fungsi *log- likelihood* diatas

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \frac{\partial \left(\sum_{i=1}^{r}\left[\sum_{j=1}^{p} \beta_j x_{ij}) - \log(\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})) \right] \right)}{\partial \beta_j}$$

$$\sum_{i=1}^{r}\left[\sum_{j=1}^{p} x_{ij} - \frac{\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj}) \sum_{j=1}^{p} x_{lj}}{\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})} \right]$$

Turunan kedua dari ln $L(\beta)$ terhadap β_j yaitu sebagai berikut:

$$\frac{\partial^2 \log L(\beta)}{\partial^2 \beta_j} = \frac{\partial y}{\partial \beta_j}\left(\frac{\partial \log L(\beta)}{\partial \beta_j} \right)$$

$$= \frac{\partial y}{\partial \beta_j}\left[\sum_{i=1}^{r}\left[\sum_{j=1}^{p} x_{ij} - \frac{\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj}) \sum_{j=1}^{p} x_{lj}}{\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum \beta_j x_{lj})} \right] \right]$$

$$= -\sum_{i=1}^{r}\left[\frac{\sum\nolimits_{l \varepsilon R(t_i)} (\sum_{j=1}^{p} x_{lj})^2 \exp(\sum_{j=1}^{p} \beta_j x_{lj})}{\sum\nolimits_{l \varepsilon R(ti)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})} - \frac{(\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj})(\sum_{j=1}^{p} x_{lj}))^2}{(\sum\nolimits_{l \varepsilon R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{lj}))^2} \right] \quad (9)$$

Persamaan diatas dapat diselesaikan secara numerik yaitu dengan iterasi menggunakan metode *Newton Raphson*.

Berikut ini adalah langkah-langkah dalam prosedur *Newton Raphson*:

Model untuk data survival tersensor biasanya menggunakan prosedur *newtonraphson* dalam memaksimalkan fungsi *partial likelihood*. Misalkan $u(\beta)$ merupakan vektor berukuran $px1$ dari turunan pertama fungsi *log-likehood* terhadap parameter β_j.

$$u(\beta)_{px1} = \begin{bmatrix} \frac{\partial \log L(\beta)}{\partial \beta_1} \\ . \\ . \\ . \\ \frac{\partial \log L(\beta)}{\partial \beta_p} \end{bmatrix}$$

Misalkan $I(\beta)$ merupakan matrik *hessian* berukuran pxp dari negatif turunan kedua fungsi *log-likelihood* yaitu dengan memisalkan: $I(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta_i \partial \beta_j}$

$$I(\beta)_{pxp} = \begin{bmatrix} \frac{\partial^2 \log L(\beta)}{(\partial \beta_1)^2} & \frac{\partial^2 \log L(\beta)}{\partial \beta_1 \beta_2} & . \; . \; . & \frac{\partial^2 \log L(\beta)}{\partial \beta_1 \beta_p} \\ \frac{\partial^2 \log L(\beta)}{\partial \beta_2 \beta_1} & \frac{\partial^2 \log L(\beta)}{\partial (\beta_2)^2} & . \; . \; . & \frac{\partial^2 \log L(\beta)}{\partial \beta_2 \beta_p} \\ . & . & & . \\ . & . & . \; . \; . & . \\ . & . & & . \\ \frac{\partial^2 \log L(\beta)}{\partial \beta_p \beta_1} & \frac{\partial^2 \log L(\beta)}{\partial \beta_p \beta_2} & . \; . \; , & \frac{\partial^2 \log L(\beta)}{\partial (\beta_p)^2} \end{bmatrix}$$

Prosedur *newton raphson* parameter $\hat{\beta}$ pada iterasi (i+1) adalah [1]

$$\hat{\beta}_{i+1} = \hat{\beta}_i + I^{-1}(\hat{\beta}_i) u(\hat{\beta}_i)$$

dengan $i = 0,1,2,\dots$ dan $I^{-1}(\hat{\beta}_i)$ merupakan invers dari $I(\hat{\beta}_i)$

Langkah iterasi dengan metode newton raphson sebagai berikut:

1. Menentukan nilai awal $\hat{\beta}_0 = 0$
2. Kemudian hitung $u(\hat{\beta}_0)$ dan $I(\hat{\beta}_0)$,sehingga diperoleh nilai β_1 dari $\hat{\beta}_1 = \hat{\beta}_0 - I^{-1}(\hat{\beta}_0)u(\hat{\beta}_0)$
3. Iterasi akan berhenti ketika perubahan fungsi *log-likelihood* cukup kecil atau apabila perubahan relatif yang terbesar dalam nilai taksir parameter yang cukup kecil.

2.10 Estimasi Parameter Pada Kejadian Bersama

Misalkan $t_1 < t_2 < \dots < t_{(r)}$ dengan r adalah banyaknya waktu kematian. Jika terdapat d_i kematian pada waktu t_i dimana *i=1,2,…r*, misalkan s_i adalah jumlah nilai variabel x_j dari semua individu yang mati pada saat t_i, dimana *j=1,2,…p* dengan $s_i = \sum_{j\epsilon D_i} x_j$, D_i himpunan individu yang mati pada saat t_i[5].

Ada beberapa cara pendekatan pada kejadian bersama, dua diantaranya yaitu sebagai berikut:

A. *Breslow Partial Likelihood*

Metode breslow mengasumsikan bahwa ukuran dari himpunan resiko pada kejadian bersama adalah sama [7]. Berikut ini adalah fungsi *partial likelihood* untuk metode *Breslow* :

Fungsi *partial likelihood* untuk metode *Breslow* :

$$L_1(\beta) = \prod_{i=1}^{r} \frac{\exp(\beta' s_i)}{[\sum_{j \varepsilon R(t_i)} \exp(\beta' x_j)]^{di}} \qquad (10)$$

Dengan d_i : banyaknya kejadian bersama pada waktu t_i

s_i : jumlah nilai variabel x_j dari semua individu yang mati pada saat t_i

$R(t_i)$:himpunan individu yang berisiko pada waktu t_i yang terdiri dari individu yang bertahan pada waktu t_i

B. *Efron Partial Likelihood*

Metode *Efron* adalah metode yang memiliki perhitungan yang lebih sederhana sama halnya dengan metode *Breslow*, tetapi metode *Efron* lebih akurat perhitungannya dari pada metode *Breslow* terutama ketika ukuran dari himpunan resiko untuk waktu kejadian bersama besar [7]. Pada pendekatan metode efron ini, himpunan resikonya diselesaikan dengan pengurangan terhadap rata-rata dari nilai fungsi hazard dari variabel ke-j, karena tidak diketahui variabel mana yang mengalami kejadian terlebih dahulu [2].

fungsi*partial likelihood* untuk metode *Efron*:

$$L_2(\beta) = \prod_{i=1}^{r} \frac{\exp(\beta' s_i)}{\prod_{j=1}^{di} \left[\sum_{k \varepsilon R(t_i)} \exp(\beta' x_k) - \frac{j-1}{di} \sum_{k \varepsilon D_i} \exp(\beta' x_k) \right]} \qquad (11)$$

denganD_i : Himpunanan semua individu yang mati pada saat t_i

x_k : variabel dari individu yang masih bertahan dan merupakan elemen dari$R(t_i)$ dan D_i

Pengujian Parameter

A. Uji *partial likelihood* rasio

Hipotesis:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Taraf Signifikansi :α

Statistik Uji:

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

Daerah Penolakan :

H$_0$ ditolak jika $G \geq x^2_{(\alpha:db=p)}$ atau p-value $\leq \alpha$

B. Uji *Wald*

Hipotesis:

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Taraf Signifikansi :α

Statistik Uji:

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

Daerah penolakan

H$_0$ ditolak jika$Z^2 \geq x^2_{(\alpha:db=p)}$ atau p-value $\leq \alpha$

Pemilihan Model terbaik

Pemilihan variabel yang masuk atau keluar dari model dapat dilakukan dengan tiga cara yaitu seleksi *forward*, eliminasi *backward*, dan prosedur *stepwise*. Prosedur seleksi *forward* atau seleksi maju dengan menambahkan variabel satu persatu kedalam model pada setiap langkahnya jika variabel memberikan penurunan nilai $-2\log L(\beta)$terbesar. Prosedur *forward*akan berhenti jika variabel berikutnya yang akan ditambahkan kedalam model tidak mengurangi nilai dari $-2\log L(\beta)$dari nilai $-2\log L(\beta)$ sebelumnya [1]. Prosedur seleksi *stepwise* merupakan kombinasi dari dua proses yaitu seleksi *forward* dan seleksi *backward*. Langkah awal seleksi *stepwise*sama dengan seleksi *forward*. Pada seleksi *stepwise* variabel yang telah ditambahkan kedalam model kemudian dibandingkan nilai $-2\log L(\beta)$, apabila tidak signifikan maka variabel dapat dihapus atau dikeluarkan dari model. Prosedur *stepwise*akan berhenti jika variabel yang masukdan keluar itu sama [3].

Pemilihan model terbaik ini menggunakan Seleksi *Backward* dengan nilai AIC terkecil. Langkah langkah seleksi *backward*:

1. Penghitungan nilai $-2\ Log\ \hat{L}$ model pertama dengan semua variabel bebas dan hitung nilai $AIC = -2\ Log\ \hat{L} + \alpha q$
2. Diambil satu variabel bebas dari model pertama untuk dikurangkan dan hitung nilai $-2\ Log\ \hat{L}$ untuk setiap variabel bebas yang telah dikurangkan dari model pertama, hitung nilai $AIC = -2\ Log\ \hat{L} + \alpha(q-1)$.
3. Nilai AIC yang terkecil pada pengecualian model yang variabel bebasnya telah dikurangkan dipilih untuk langkah berikutnya.
4. Diambil dua variabel bebas dari model pertama untuk dikurangkan dan hitung nilai $-2\ Log\ \hat{L}$ untuk setiap pengecualian model dengan variabel bebas yang telah dikurangkan dari model pertama, hitung nilai $AIC = -2\ Log\ \hat{L} + \alpha(q-2)$. Nilai AIC yang terkecil yang dipilih untuk langkah selanjutnya.
5. Apabila setelah pengurangan variabel bebas perbandingan nilai $-2\ Log\ \hat{L}$ antar model tidak berbeda signifikan, dan memperoleh nilai AIC yang terkecil maka proses pengurangan variabel bebas pada model dihentikan.

3. Hasil dan Pembahasan

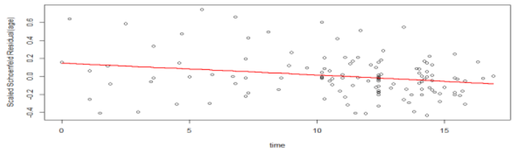
3.1 Pengujian asumsi proportional hazard

Asumsi *proportional hazard* berarti bahwa rasio fungsi *hazard* dari dua individu konstan dari waktu atau ekuivalen dengan pernyataan bahwa fungsi *hazard* suatu individu terhadap fungsi *hazard* individu yang lain adalah *proportional*. Pengujian asumsi ini menggunakan *scaled schoenfeld residual*. Hasil uji asumsi menggunakan *scaled schoenfeld residual* dengan menggunakan program R.

Berikut ini adalah plot *scaled schoenfeld residual* dari masing-masing variabel bebas yang terdapat pada kasus ketahanan hidup pasien diabetes :

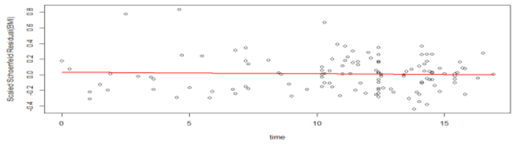
 ISSN : 2085-9902

A. Variabel Umur



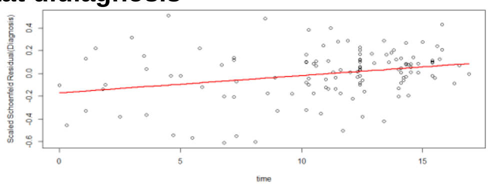
Gambar 1. Plot *Scaled Schoenfeld Residual* untuk Variabel Umur

B. Variabel BMI



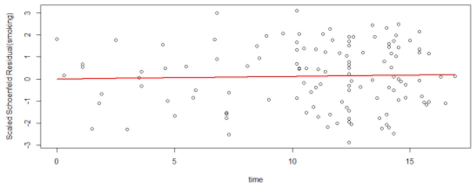
Gambar 2. Plot *Scaled Schoenfeld Residual* untuk Variabel BMI

C. Variabel Umur saat didiagnosis



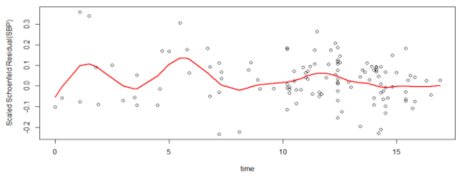
Gambar 3. Plot *Scaled Schoenfeld Residual* untuk Variabel Diagnosis

D. Variabel Status Merokok



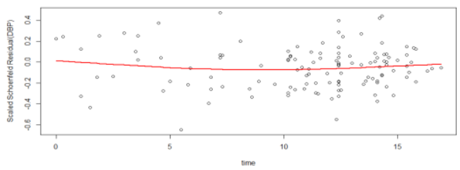
Gambar 4. Plot *Scaled Schoenfeld Residual* untuk Variabel Status Merokok

E. Variabel SBP



Gambar 5. Plot *Scaled Schoenfeld Residual* untuk Variabel SBP

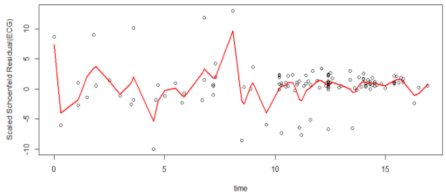
F. Variabel DBP



Gambar 6. Plot *Scaled Schoenfeld Residual* untuk Variabel DBP

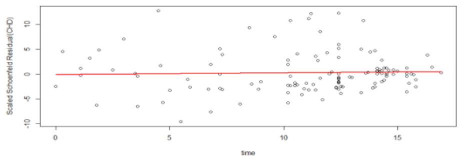
 ISSN : 2085-9902

G. Variabel ECG



Gambar 7. Plot *Scaled Schoenfeld Residual* untuk Variabel ECG

H. Variabel CHD



Gambar 8. Plot *Scaled Schoenfeld Residual* untuk Variabel CHD

Berdasarkan pengujian asumsi *proportional hazard* dengan menggunakan pendekatan *scaled schoenfeld residual* dapat disimpulkan bahwa terdapat 6 variabel yang memenuhi asumsi proportional hazard yaitu varibel umur, BMI, diagnosis, status merokok, DBP, CHD. Sementara variabel SBP dan ECG tidak memenuhi asumsi *proportional hazard*, sehingga kedua variabel tersebut tidak perlu diestimasi dan dapat langsung dikeluarkan dari model.

Estimasi *Efron Partial Likelihood*

Tabel 1. Hasil Estimasi parameter *Efron Partial Likelihood*

Variabel	Coef	Se(coef)	Z	Pr(>|z|)
Umur	0,00906	0,01777	0,510	0,61006
BMI	0,01345	0,01647	0,817	0,41397
Diagnosis	0,00047	0,01708	0,028	0,97781
Satus merokok	0,17400	0,11834	1,470	0,14147
DBP	-0,01402	0,00939	-1,492	0,13573
CHD	0,83334	0,23642	3,525	0,00042

Berdasarkan hasil estimasi diatas, diasumsikan semua variabel berpengaruh dalam model sehingga diperoleh model awal sebagai berikut:

$$h(t,X) = h_0(t)\exp(0{,}00906\,X_1 + 0{,}01345\,X_2 + 0{,}00047\,X_3 + 0{,}17400\,X_4 - 0{,}01402\,X_5 + 0{,}83334\,X_6$$

Untuk menguji semua variabel berpengaruh atau tidak dalam model maka dapat diuji dengan Uji *partial likelihood* rasio.Dari hasil output menggunakan software R diperoleh nilai G yaitu sebagai berikut:$G = -2[-488{,}8642 - (-478{,}2524)] = 21{,}2236$, karena nilai G = 21,2236 dan $\geq \mathcal{X}^2_{(0{,}05:6)} = 12{,}5916$ dan nilai *p-value* = 0,001672327 <0,05 sehingga H_0 ditolak dan dapat disimpulkan bahwa semua variabel berpengaruh terhadap model.

Estimasi *Breslow Partial Likelihood*

Tabel 2. Hasil Estimasi Parameter *Breslow Partial Likelihood*

Variabel	Coef	Se(coef)	Z	Pr(>|z|)
Umur	0,01136	0,01778	0,639	0,52298
BMI	0,01313	0,01647	0,798	0,42504
Diagnosis	- 0,00189	0,01711	-0,111	0,91200
Status merokok	0,16589	0,11816	0,404	0,16034
DBP	- 0,01385	0,00935	-1,481	0,13855
CHD	0,80811	0,23469	3,443	0,00057

Berdasarkan hasil estimasi diatas, diasumsikan semua variabel berpengaruh dalam model sehingga diperoleh model awal sebagai berikut:

$h(t,X) = h_0(t)\exp(0{,}01136X_1 + 0{,}01313X_2 - 0{,}00189X_3 +$
$0{,}16589X_4 - 0{,}01385X_5 + 0{,}80811X_6$

Untuk menguji semua variabel berpengaruh atau tidak dalam model maka dapat diuji dengan Uji *partial likelihood* rasio.Dari hasil output menggunakan software R diperoleh nilai G yaitu sebagai berikut:
$G = -2[-492{,}1838 - (-481{,}8810)] = 20{,}6056$karena nilai G = 20,6056 dan $\geq \mathcal{X}^2_{(0{,}05:6)} = 12{,}5916$ dan nilai *p-value* = 0,002159074 <0,05 sehingga H_0 ditolak dan dapat disimpulkan bahwa semua variabel berpengaruh terhadap model.

Pemilihan Model Terbaik

Tabel 3. Hasil Simulasi Pemilihan Model Terbaik *Efron* dengan Metode *Backward*

Variabel	AIC
Semua Variabel	968,5
Semua variabel tanpa diagnosis	966,51
Semua variabel tanpa diagnosis dan BMI	965,23
Variabel smoking,DBP dan CHD	963,73
Variabel DBP dan CHD	963,26
Variabel CHD	963

Berdasarkan hasil yang diperoleh dari Tabel 3 dapat disimpulkan bahwa model terbaik yang diperoleh dari estimasi efron dengan seleksi *backward* berdasarkan kriteria nilai AIC terkecil maka model *cox proportional hazard* yang terbaik yaitu model yang hanya terdiri dari variabel bebas CHD berarti variabel ini memiliki pengaruh yang signifikan terhadap waktu survival ketahanan hidup pasien diabetes.

Tabel 4. Hasil Simulasi Pemilihan Model Terbaik *Breslow* dengan Metode *Backward*

Variabel	AIC
Semua Variabel	975,76
Semua variabel tanpa diagnosis	973,77
Semua variabel tanpa diagnosis dan BMI	972,41
Variabel smoking,DBP dan CHD	970,98
Variabel DBP dan CHD	970,27
Variabel CHD	969,99

Berdasarkan hasil yang diperoleh dari Tabel 4dapat disimpulkan bahwa model terbaik yang diperoleh dari estimasi breslow dengan seleksi *backward* berdasarkan kriteria nilai AIC terkecil maka model *cox proportional hazard* yang terbaik yaitu model yang hanya terdiri dari variabel bebas CHD berarti variabel ini memiliki pengaruh yang signifikan terhadap wakt\
survival ketahanan hidup pasien diabetes.

Perbandingan Model terbaik dengan pendekatan *efron* dan *breslow*

Tabel 5. Hasil analisis perbandingan model *efron* dan *breslow*

	AIC
Efron	963
Breslow	969,99

Berdasarkan hasil analisis yang diperoleh dari Tabel diatas, dapat dilihat bahwa nilai AIC terkecil ada pada model *efron* yaitu dengan AIC=963, sehingga dapat disimpulkan berdasarkan kriteria AIC terkecil maka model terbaik yang diperoleh yaitu model dengan pendekatan estimasi *efron partial likelihood* dengan model yang hanya terdiri dari variabel CHD.

Tabel 6. Pengujian Parameter pada model regresi *cox proportional hazard* secara *partial* dan keseluruhan

Variabel	Coef	Exp(coef)	Se(coef)	Z	P
CHD	0,879	2,409	0,2045	4,3	1,7 e-05

Berdasarkan Tabel 6 dapat disimpulkan sebagai berikut:
Variabel CHD berpengaruh terhadap waktu survival, hal tersebut dapat dilihat dari nilai Z = 18,49 ≥ $\mathcal{X}^2_{(0,05:1)}$ = 3,84 atau *p-value* dari uji *wald* pada tabel diatas yaitu 1,7 e-05 < α = 0,05 maka H_0 ditolak, dapat disimpulkan bahwa variabel CHD memiliki pengaruh yang signifikan terhadap model atau dapat dikatakan bahwa variabel CHD berpengaruh terhadap ketahanan hidup pasien diabetes.

Apabila dilihat dari uji *partial likelihood* rasio maka diperoleh hasil sebagai berikut:
nilai G = 16,72512 dan ≥ $\mathcal{X}^2_{(0,005:1)} = 3,84$ dan nilai *p-value* = 4,321e-05 <0,05 sehingga H_0 ditolak dan dapat disimpulkan bahwa variabel CHD berpengaruh terhadap model.

Berdasarkan hasil estimasi parameter dan pengujian parameter diatas maka diperolehlah model regresi *cox proportional hazard* yaitu sebagai berikut:

$$h(t, X) = h_0(t)\exp(0{,}8794X_6)$$

4. Kesimpulan

Berdasarkan hasil pembahasan mengenai regresi *cox proportional hazard* untuk kejadian bersama pada data ketahanan hidup pasien diabetes, maka model *cox* yang diperoleh yaitu model *cox* yang hanya terdiri dari variabel CHD dengan mengunakan estimasi *Efron Partial Likelihood* dan dengan menggunakan seleksi *backward* dilihat dari nilai AIC yang terkecil. Model *cox proportionalhazard*. untuk penelitian selanjutnya dapat menggunakan 3 perbandingan metode dalam mengestimasi parameter yaitu membandingkan antara estimasi *efron partiallikelihood* dengan metode *exact partial likelihood*, estimasi breslow *partiallikelihood* dengan metode estimasi *exact partial likelihood* atau dapat juga dengan cara membandingkan ketiga estimasi pada data *ties* tersebut agar memperoleh hasil yang lebih baik lagi.

Referensi

[1] Collet,D. "*Modeling Survival Data In Medical Research".* Chapman and Hall, USA. 2003.
[2] Feriana, Dwi, Anjar. "Model Cox Stratifikasi".*Skripsi Universitas Indonesia*, Jakarta, 2011.
[3] Hosmer, D. W, dkk. "*Applied Survival Analysis:Regression Modeling Of Time To Event Data".* Jhon Wiley, New Jersey. 2008.
[4] Iskandar, Muhammad. "Model Cox Proportional Hazard pada kejadian bersama".*Skripsi Universitas Negeri Yogyakarta*, Yogyakarta, 2014.
[5] Klein, J. P dan Moeschberger, M. L. "*Survival Analysis* :*Techiques for Censored and Truncated Data Second Editio*n". Springer-Verlag, New York. 2003.
[6] Lee, E. T dan Wang, J. W. "*Statistical Methods for Survival Data Analysis Third Edition*". John Wiley & Sons, Inc, New Jersey. 2003.
[7] Xinxin. "A Study Of Ties And Time Varying Covariates In Cox Proportional Hazard Model". *Thesis The University Of Guelph*, 2011.