

Diajukan kepada Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta
untuk Memenuhi Sebagian Persyaratan guna Memperoleh Gelar Sarjana Sains



Oleh:
Anna Sandi Rahayu
NIM. 10305141028

PROGRAM STUDI MATEMATIKA
JURUSAN PENDIDIKAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI YOGYAKARTA
2015

i

BAB I PENDAHULUAN

A. Latar Belakang

Analisis survival merupakan salah satu prosedur dalam statistika untuk menganalisis data dimana variabel yang diperhatikan adalah waktu sampai terjadinya suatu kejadian dan variabel-variabel lain yang diduga mempengaruhi waktu survival (Kleinbaum dan Klein, 2005: 4). Kejadian yang diamati berupa kematian akibat suatu penyakit, kecelakaan, kelulusan sekolah, atau kerusakan suatu alat. Waktu sampai terjadinya suatu kejadian tersebut dikenal dengan istilah waktu survival. Waktu terjadinya suatu kejadian dapat berupa hari, bulan, ataupun tahun dari awal penelitian pada suatu subyek sampai suatu kejadian terjadi pada subyek tersebut. Data tentang pengamatan jangka waktu dari awal pengamatan sampai terjadinya suatu peristiwa disebut data survival.

Salah satu yang membedakan analisis survival dengan prosedur analisis statistika yang lain yaitu terdapat konsep pensensoran (Kleinbaum dan Klein, 2005: 5). Dalam mendapatkan data survival seringkali dijumpai subyek yang tidak mengalami kejadian sampai batas waktu pengamatan. Untuk memperoleh data yang lengkap sampai semua subyek mengalami kejadian membutuhkan waktu yang lama sehingga pengamatan yang dilakukan kurang efektif dan menghabiskan biaya yang besar. Untuk mengatasi hal itu maka dilakukan pensensoran data.

1

ANALISIS SURVIVAL UNTUK DATA KEJADIAN BERULANG TIDAK IDENTIK DENGAN MODEL COX STRATIFIKASI PWP-GAP TIME

Oleh:
Anna Sandi Rahayu
NIM. 10305141028

ABSTRAK

Model Cox Stratifikasi PWP-Gap Time adalah model yang dikemukakan oleh Prentice, William dan Peterson untuk menganalisis data survival kejadian berulang tidak identik. Penulisan skripsi ini bertujuan untuk menjelaskan prosedur pembentukan model Cox Stratifikasi PWP-Gap Time pada data kejadian berulang tidak identik dan penerapannya untuk mengetahui faktor-faktor yang mempengaruhi kasus kekambuhan pada pasien penderita stroke.

Estimasi parameter model Cox Stratifikasi PWP-Gap Time menggunakan metode *Maximum Partial Likelihood Estimation* yaitu dengan memaksimalkan fungsi likelihood stratifikasi, yang diperoleh dengan mengalikan fungsi *partial likelihood* dari masing-masing subyek pada masing-masing strata. Estimasi parameter model Cox Stratifikasi PWP-Gap Time pada masing-masing strata dibantu dengan perangkat lunak R 3.1.0.

Prosedur pembentukan model Cox Stratifikasi PWP-Gap time untuk analisis survival kejadian berulang tidak identik terdiri dari 10 tahap, yaitu (1) penghitungan waktu survival *gap time*. (2) Penyusunan data kejadian berulang. (3) Pensaksiran parameter. (4) Pemilihan model terbaik dengan *forward procedure*. (5) Pengujian parameter dengan menggunakan uji *Wald*. (6) Pengujian asumsi proporsional hazard. (7) Stratifikasi variabel ulangan sesuai dengan banyaknya kejadian yang dialami subyek penelitian. (8) Pensaksiran parameter model Cox Stratifikasi PWP-Gap Time pada masing-masing strata. (9) Pengujian asumsi model interaksi Cox Stratifikasi PWP-Gap Time dan (10) interpretasi model. Data kejadian berulang tidak identik 653 pasien penderita stroke dianalisis menggunakan model Cox Stratifikasi PWP-Gap Time, kemudian diperoleh variabel yang berpengaruh secara signifikan pada kekambuhan penyakit stroke adalah variabel kebiasaan merokok. Pasien stroke yang memiliki kebiasaan merokok berpeluang mengalami kekambuhan 1,6 kali dari pasien yang tidak memiliki kebiasaan merokok.

Kata Kunci: PWP-Gap Time, conditional, kejadian berulang, cox stratifikasi, strata, survival, stroke

vii

Menurut David Collet (2003: 2) dalam analisis survival terdapat 3 jenis pensensoran yaitu sensor kiri, sensor kanan, dan sensor interval.

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*. Regresi *Cox Proportional Hazard* mempunyai variabel terikat berupa waktu survival dan variabel bebas yaitu variabel yang diduga berpengaruh terhadap waktu survival. Dalam Regresi *Cox Proportional Hazard* terdapat asumsi proporsional hazard yang harus dipenuhi yaitu bahwa fungsi hazard suatu individu terhadap fungsi hazard individu yang lain adalah proporsional (Guo, 2010: 77).

Analisis survival sebenarnya tidak hanya terbatas pada analisis waktu yang diperlukan suatu subyek hingga mengalami suatu kejadian tetapi analisis survival juga dapat digunakan untuk melakukan analisis kejadian yang berulang, baik kejadian berulang identik maupun kejadian berulang tidak identik. Sebagai contoh kejadian berulang identik adalah kasus serangan jantung, baik serangan jantung pertama, kedua, dan seterusnya dianggap sama dan tidak dibedakan tipenya. Contoh untuk kejadian berulang tidak identik adalah kekambuhan penyakit stroke dan kasus *ocular degeneration* (kemunduran ketajaman visual) karena kejadian berulang tersebut mengindikasikan tahapan yang lebih parah dari sebelumnya. Dalam bidang kedokteran, seorang pasien dengan penyakit tertentu mungkin saja mengalami kekambuhan beberapa kali. Setiap kali terjadi kambuh dapat diindikasikan gejala yang lebih parah dari sebelumnya. Kejadian berulang yang

2

Proportional Hazard untuk mengatasi variabel bebas yang tidak memenuhi asumsi proporsional hazard. Apabila dalam model *Cox Proportional Hazard* ditemukan variabel bebas yang tidak memenuhi asumsi proporsional hazard maka dilakukan modifikasi pada model Cox tersebut. Modifikasi dilakukan dengan menstratifikasi variabel yang tidak memenuhi asumsi proporsional hazard menjadi strata. Variabel bebas yang tidak memenuhi asumsi proporsional hazard tidak masuk kedalam model Cox Stratifikasi.

Model Cox Stratifikasi PWP-Gap Time merupakan model yang digunakan untuk analisis kejadian berulang tidak identik yang dikemukakan oleh Prentice, William dan Peterson (1981). Model Cox Stratifikasi PWP-Gap Time ini mengasumsikan suatu subyek tidak berisiko mengalami kejadian berikutnya sebelum mengalami kejadian sebelumnya, yang lebih dikenal dengan kejadian *conditional*. Model Cox Stratifikasi PWP-Gap Time memperhatikan waktu survival antara 2 kejadian yang berurutan dalam menentukan himpunan subyek yang berisiko untuk mengalami kejadian berulang (Hosmer dan Lemeshow, 2008: 307). Waktu survival *gap time* yang digunakan diperoleh dengan mencari selisih waktu suatu subyek mengalami kejadian dengan waktu subyek tersebut mengalami kejadian berikutnya. Penaksiran parameter model Cox Stratifikasi PWP-Gap Time dilakukan untuk mengetahui pengaruh variabel bebas yang

3

kasus stroke yang berat dapat terjadi kematian, pada kasus yang dapat bertahan hidup beberapa kemungkinan terjadi stroke berulang. Diperkirakan setiap tahunnya sebanyak 500.000 penduduk Indonesia terkena serangan stroke, sekitar 2,5 % atau 125.000 orang meninggal, dan sisanya cacat ringan maupun berat. Secara umum, dapat dikatakan angka kejadian stroke adalah 200 per 100.000 penduduk. Menurut Yayasan Stroke Indonesia (2012: 1), faktor risiko stroke, yaitu kondisi seseorang yang memudahkan orang tersebut mengalami serangan stroke pada suatu saat, antara lain jenis kelamin, kegemukan (*obesity*), riwayat diabetes, tekanan darah (*blood pressure*), kebiasaan merokok, dan kadar kolesterol dalam darah.

B. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalahnya adalah sebagai berikut:

1. Bagaimana prosedur pembentukan model Cox Stratifikasi PWP-Gap Time untuk data kejadian berulang tidak identik?
2. Bagaimana penerapan model Cox Stratifikasi PWP-Gap Time pada kasus kekambuhan penyakit stroke?

C. Tujuan Penulisan

Sesuai dengan rumusan masalah, maka tujuan dari penulisan skripsi ini adalah sebagai berikut:

1. Menjelaskan prosedur pembentukan model Cox Stratifikasi PWP-Gap Time untuk data kejadian berulang tidak identik.
2. Menjelaskan penerapan model Cox Stratifikasi PWP-Gap Time untuk kasus kekambuhan penyakit stroke.

5

diduga berpengaruh terhadap waktu survival *gap time* maka dilakukan. Penaksiran parameter model Cox Stratifikasi PWP-Gap Time dibantu dengan perangkat lunak R 3.1.0.

Beberapa artikel yang mengkaji tentang Model Cox Stratifikasi PWP-Gap Time telah ditulis oleh beberapa orang antara lain Hyun Ja Lim (2007: 290-299) yang membahas model Cox Stratifikasi yang dapat digunakan untuk analisis data kejadian berulang. Kelly dan Lim (2000: 13-33) menyatakan bahwa model Cox Stratifikasi PWP-Gap Time merupakan model yang digunakan untuk menganalisis kasus penyakit infeksi pada anak-anak. Pada kasus penyakit infeksi pada anak-anak tersebut model PWP-Gap time digunakan untuk menentukan apakah *treatment* pemberian vitamin A efektif untuk mencegah infeksi dengan memperhatikan waktu yang dihitung sejak dari kejadian infeksi sebelumnya. Selanjutnya model Cox Stratifikasi telah dibahas oleh Dwi Anjar Feriana (2011) dalam skripsinya untuk menganalisis faktor-faktor yang berpengaruh terhadap waktu saat penderita penyakit paru-paru menjalani pengobatan hingga sembuh ataupun meninggal.

Penerapan model Cox Stratifikasi PWP-Gap Time dapat dilakukan pada kasus kekambuhan penyakit stroke. Stroke merupakan gangguan fungsi saraf pusat yang berkembang sangat cepat, baik menit maupun jam, dengan perburukan ringan sampai berat kemudian menetap atau bahkan membaik secara cepat atau perlahan, tergantung tingkat keparahan stroke dan cepat serta tepatnya intervensi pengobatan. Perjalanan penyakit stroke beragam, penderita stroke dapat pulih sempurna, ada pula yang cacat ringan, cacat sedang sampai cacat yang berat. Pada

4

D. Manfaat Penulisan

Penulisan skripsi ini diharapkan dapat memberikan manfaat antara lain:

1. Bagi mahasiswa

Pengembangan ilmu teoritis yang dipelajari diperkuliahan dan penambahan wawasan analisis survival khususnya model Cox Stratifikasi PWP-Gap Time dengan waktu survival berupa *gap time* untuk data kejadian berulang tidak identik.

2. Bagi Perpustakaan Jurusan Pendidikan Matematika UNY

Menambah referensi mengenai model Cox Stratifikasi PWP-Gap Time dengan waktu survival berupa *gap time* untuk data kejadian berulang tidak identik bagi mahasiswa Jurusan Pendidikan Matematika.

3. Bagi Pembaca

Penelitian ini dapat dijadikan referensi pembaca untuk mengembangkan model Cox Stratifikasi PWP-Gap Time pada kasus-kasus kejadian berulang tidak identik yang terjadi di Indonesia.

6

Analisis survival merupakan teknik analisis yang perhatian utamanya terletak pada waktu. Dalam analisis survival, faktor yang mempengaruhi ketahanan hidup suatu subyek ditentukan pada awal penelitian. Selama penelitian tidak ada faktor tambahan lain yang terjadi, yang diluar kendali peneliti. Asumsi-asumsi lain meliputi (Gudono, 2011: 318) :

- Masing-masing observasi atau subyek individu dalam sampel bersifat independen.
- Masing-masing kejadian (*event*) bersifat *mutually exclusive* saling meniadakan dan bersifat *exhaustive* (menyeluruh atau berisi semua kemungkinan). Maksudnya individu tidak boleh berada dalam lebih dari satu kondisi. Contohnya, "mati", "hidup", "bangkrut" adalah kondisi-kondisi yang saling meniadakan (*mutually exclusive*) karena tidak mungkin suatu individu pada saat yang bersamaan mengalami "mati" dan "hidup". Sekaligus bersifat *exhaustive* (semua kemungkinan sudah tercakup) karena hanya ada kemungkinan mati atau hidup.
- Semua individu terbebas dari semua kejadian (*event*) pada saat mereka memasuki periode penelitian.
- Penyensoran bersifat independen artinya penyensoran tidak berhubungan dengan peluang terjadinya kejadian (*event*) yang diamati.

7

Misalkan peneliti sedang menyelidiki penyakit baru yaitu penyakit "X" yang berbahaya. Periode amatan dilakukan peneliti selama 5 tahun pada masing-masing subyek. Terdapat 3 pasien yang sudah terserang penyakit tersebut dan diambil sebagai subyek penelitian. Subyek (1) mulai diamati pada tahun 2000 dan penelitian berakhir pada tahun 2005. Subyek (1) meninggal tahun 2007 dan penelitian berakhir pada tahun 2005 sehingga pada tahun 2006 dan 2007 tidak dilakukan pengamatan pada subyek ini. Untuk subyek (2) mulai diamati pada tahun 2004 dan meninggal pada tahun 2007 ketika masih dalam periode amatan. Subyek (3) mulai diamati 2005 dan meninggal pada tahun 2010, sementara peneliti memutuskan untuk menghentikan amatan pada tahun 2008, 2 tahun lebih awal dari periode amatan yang telah direncanakan peneliti.

Dalam konsep analisis survival waktu dapat diatur (seolah-olah) mulai (*start*) pada saat bersamaan seperti pada gambar (b) sesuai konsep waktu survival yang terdiri dari waktu mulai dan waktu akhir. Menurut *calender time*, ketiga subyek mulai diamati pada waktu yang berbeda-beda (Gudono, 2011: 322). Penentuan waktu awal dan akhir dalam analisis survival sangat penting untuk menentukan siapa saja subyek yang berisiko untuk suatu kejadian.

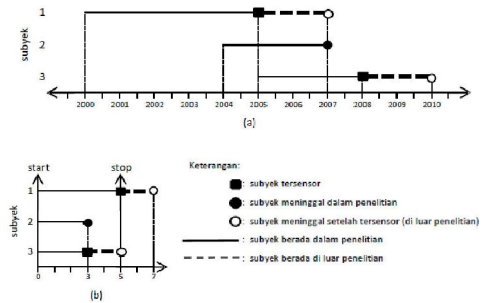
2. Penyensoran

Data yang digunakan dalam analisis survival harus memiliki beberapa unsur yang harus dipenuhi. (1) titik asal ketika dimulainya pengamatan terhadap suatu subyek yang harus didefinisikan dengan tepat. (2) skala waktu dan (3) kejadian gagal (*failure event*) yang juga harus didefinisikan dengan tepat. Data

9

1. Waktu Survival (*Survival Time*)

Analisis survival merupakan analisis mengenai waktu sampai terjadinya suatu peristiwa atau kejadian atau waktu antara dua kejadian. Permulaan (*start*) dan akhir (*end*) interval waktu perlu ditetapkan oleh peneliti. Interval waktu selama subyek diamati merupakan periode amatan (*observation period*). Penentuan awal dan akhir waktu survival dapat bervariasi dan penting untuk menentukan siapa saja subyek yang berisiko (*at risk*) untuk suatu kejadian. Menurut Gudono (2011: 320), *survival time* berbeda dengan *calender time*, *survival time* diukur dan ditetapkan berdasarkan kesesuaiannya dengan mulainya peristiwa tertentu. Secara visual *calender time* dan *survival time* seperti pada gambar 2.1 (a) dan (b),



Gambar 2.1 *Calender time* (a) dan *Survival time* (b)

8

tersensor merupakan data yang tidak dapat diamati secara utuh karena subyek penelitian hilang ataupun dengan alasan lain sehingga tidak dapat diambil datanya, atau sampai akhir penelitian subyek tersebut belum mengalami kejadian tertentu. Sedangkan data yang berada dalam keadaan sebaliknya disebut data yang tidak tersensor (Lee dan Wang, 2003: 2). Menurut Kleinbaum dan Klein (2005: 6) data dikatakan tersensor apabila:

- Subyek tidak mengalami kejadian selama penelitian.
- Subyek tidak mengikuti *treatment* yang diberikan sampai penelitian berakhir (*lost to follow up*).
- Subyek terpaksa dihentikan dari penelitian karena meninggal dimana meninggal bukan peristiwa yang diperhatikan peneliti atau alasan lain (misal reaksi pengobatan yang merugikan atau risiko lain).

Pengamatan dilakukan untuk memperoleh data survival yang terpercayaa dan tepat. Namun untuk memperoleh data yang lengkap sampai subyek mengalami kejadian yang diinginkan, peneliti membutuhkan waktu yang lama dan seringkali menelan banyak biaya. Untuk mengatasi permasalahan tersebut dilakukan penyensoran. Dengan melakukan penyensoran, data yang diperoleh tetap lengkap dengan rentang waktu pengamatan yang telah ditetapkan untuk observasi subyek. Konsep penyensoran inilah yang membedakan analisis survival dengan analisis bidang statistik yang lain (Kleinbaum dan Klein, 2005).

Menurut David Collet (2003: 2) dalam analisis survival terdapat 3 jenis penyensoran yaitu:

10

seorang yang positif menderita HIV. Peneliti mencatat kejadian tepatnya seseorang tersebut mendapatkan tes pertamanya dan positif HIV namun peneliti tidak memiliki catatan tentang waktu tepatnya seseorang tersebut terjangkit virus pertama HIV dan kapan tepatnya virus itu berkembang. Dengan demikian penderita HIV tersebut tersensor kiri yaitu ketika mengalami kejadian tes pertama dengan hasil positif menderita HIV.

b. Sensor kanan (*right censoring*)

Sensor kanan terjadi ketika subyek yang masuk dalam observasi dapat diamati secara penuh namun hingga akhir penelitian belum mengalami kejadian. Sebagai contoh suatu subyek penderita kanker diteliti sejak umur 10 tahun saat mulai didiagnosa dokter subyek terkena kanker. Peneliti menetapkan periode amatan selama 5 tahun. Namun subyek pindah ke negara lain saat berumur 14 tahun. Subyek masih memiliki waktu survival dalam penelitian setidaknya satu tahun. Sehingga waktu pengamatan individu tersebut dikatakan tersensor kanan.

c. Sensor interval (*interval censoring*)

Sensor interval adalah sensor yang waktu survival berada dalam suatu selang tertentu. Sebagai contohnya, jika catatan medis menunjukkan bahwa pada saat berumur 30 tahun penderita HIV dalam contoh diatas dalam kondisi sehat,

2. Fungsi Survival

Fungsi survival dinotasikan $S(t)$ didefinisikan sebagai peluang suatu individu dapat bertahan hidup dengan waktu survival sama dengan atau lebih dari waktu t (Lawless, 1982: 8).

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x) dx. \quad (2.4)$$

Menggunakan definisi fungsi distribusi kumulatif dari T , fungsi survival dapat dinyatakan sebagai berikut:

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(T \leq t) \\ &= 1 - F(t) \end{aligned}$$

maka diperoleh,

$$\begin{aligned} F(t) &= 1 - S(t) \\ \frac{d(F(t))}{dt} &= \frac{d(1 - S(t))}{dt} \\ F'(t) &= -S'(t) \\ f(t) &= -S'(t). \end{aligned} \quad (2.5)$$

Diperoleh hubungan antara fungsi kepadatan peluang, fungsi distribusi kumulatif dari T dengan fungsi survival yaitu,

$$f(t) = F'(t) = -S'(t). \quad (2.6)$$

3. Fungsi Hazard

Fungsi ini juga dikenal sebagai *hazard rate* yang dinotasikan dengan $h(t)$. Fungsi hazard didefinisikan sebagai kelajuan suatu individu untuk mengalami kejadian dalam interval waktu dari t sampai $t + \Delta t$ jika diketahui individu

belum terjangkit virus HIV. Katakan penderita melakukan tes pertama saat berumur 40 tahun. Dengan demikian usia saat didiagnosis positif HIV adalah antara 30 dan 40 tahun.

B. Dasar Teori Analisis Survival

Misalkan T suatu variabel acak positif yang menunjukkan waktu survival dari masing-masing subyek. T menunjukkan waktu, sehingga nilai-nilai yang mungkin untuk T adalah $T \geq 0$. t merupakan nilai tertentu dari variabel acak T . Distribusi dari variabel acak T dapat dinyatakan melalui 3 cara yaitu:

1. Fungsi Kepadatan Peluang

Fungsi kepadatan peluang atau *Probability Density Function (PDF)* merupakan peluang satu subyek mati atau gagal dalam interval waktu t sampai $t + \Delta t$. Fungsi kepadatan peluang dinotasikan dengan $f(t)$ dan dirumuskan sebagai berikut (Lawless, 1982: 8).

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t < T < (t + \Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right]. \quad (2.1)$$

Jika T merupakan variabel acak non negatif maka $F(t)$ merupakan fungsi distribusi kumulatif dari T . $F(t)$ didefinisikan sebagai peluang suatu individu mengalami peristiwa gagal pada waktu t ditulis berikut:

$$F(t) = P(T \leq t) = \int_0^t f(x) dx. \quad (2.2)$$

Berdasarkan persamaan (2.2) diperoleh,

$$f(t) = \frac{d(F(t))}{dt} = F'(t). \quad (2.3)$$

tersebut masih dapat bertahan hidup sampai dengan waktu t . Secara matematis dinyatakan sebagai berikut (Lawless, 1982: 8).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (2.7)$$

Berdasarkan teori peluang, bahwa peluang kejadian A dengan syarat kejadian B yaitu:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.8)$$

dan T merupakan variabel acak maka dari persamaan (2.7) diperoleh,

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, T \geq t)}{P(T \geq t), \Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{S(t), \Delta t} \\ &= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \\ &= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t) - P(T < t)}{\Delta t} \\ &= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \\ h(t) &= \frac{f(t)}{S(t)}. \end{aligned} \quad (2.9)$$

Kemudian persamaan (2.5) disubstitusikan ke dalam persamaan (2.9) sehingga diperoleh (Lawless, 1982: 9):

$$h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)}$$

sehingga dari persamaan diatas diperoleh,

$$\begin{aligned} \int_0^t h(x)dx &= - \int_0^t \frac{d \log S(x)}{dx} dx \\ - \int_0^t h(x)dx &= \int_0^t \frac{d \log S(x)}{dx} dx \\ - \int_0^t h(x)dx &= \log S(x) \Big|_0^t \\ - \int_0^t h(x)dx &= \log S(t) - \log S(0). \end{aligned}$$

Diketahui bahwa $S(0) = 1$ dan $\log S(0) = 0$, sehingga diperoleh,

$$\begin{aligned} - \int_0^t h(x)dx &= \log S(t) \\ S(t) &= \exp \left[- \int_0^t h(x)dx \right]. \end{aligned} \quad (2.10)$$

Berdasarkan fungsi hazard yang diperoleh dari persamaan (2.10) maka fungsi kumulatif hazard dinyatakan dengan $H(t)$ (Lawless, 2007: 9) yaitu sebagai berikut,

15

bergantung pada nilai x_1, x_2, \dots, x_p dari p variabel bebas X_1, X_2, \dots, X_p . Himpunan nilai variabel bebas pada model Cox direpresentasikan oleh vektor \mathbf{x} , sehingga $\mathbf{x} = (x_1, x_2, \dots, x_p)$. Diasumsikan $\mathbf{X} = X_1, X_2, \dots, X_p$ merupakan variabel bebas yang independen terhadap waktu. Model Cox dapat dituliskan sebagai berikut.

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.13)$$

dengan memisalkan,

$h_0(t)$ = fungsi dasar hazard
 $\beta_1, \beta_2, \dots, \beta_p$ = parameter regresi,
 x_1, x_2, \dots, x_p = nilai dari variabel bebas X_1, X_2, \dots, X_p .

Rumus model Cox pada persamaan (2.13) memiliki sifat bahwa jika semua X sama dengan nol, maka rumus tereduksi menjadi fungsi hazard dasar $h_0(t)$. Dengan demikian $h_0(t)$ dianggap sebagai awal atau dasar dari fungsi hazard, dapat dituliskan sebagai berikut.

$$\begin{aligned} h(t, \mathbf{x}) &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \\ &= h_0(t) \exp(\beta_1 \times 0 + \beta_2 \times 0 + \dots + \beta_p \times 0) \\ &= h_0(t) \exp(0) \\ &= h_0(t)(1) \\ h(t, \mathbf{x}) &= h_0(t). \end{aligned} \quad (2.14)$$

Persamaan (2.13) dapat dituliskan dalam bentuk sebagai berikut,

$$\begin{aligned} \log h(t, \mathbf{x}) &= \log(h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \\ \log \frac{h(t, \mathbf{x})}{h_0(t)} &= \log(\exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \end{aligned}$$

17

$$H(t) = \int_0^t h(x)dx, \quad (2.11)$$

Selain itu, persamaan (2.10) dapat dituliskan sebagai berikut (Lawless, 2007).

$$S(t) = \exp[-H(t)], \quad (2.12)$$

C. Model Cox Proportional Hazard

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 *Cox Proportional Hazard*. Model *Cox Proportional Hazard* memiliki asumsi *proportional hazard* yaitu asumsi yang menyatakan bahwa fungsi hazard dari individu yang berlainan adalah proporsional atau rasio dari fungsi hazard dua individu yang berlainan adalah konstan (Guo, 2010: 77). Terdapat model parametrik dan model semiparametrik dalam analisis survival. Model parametrik dalam analisis survival antara lain model Weibull dan model Gamma, dengan waktu survival berturut-turut berdistribusi *weibull* dan *gamma* (Kleinbaum & Klein, 2005: 357). Model Cox merupakan model berdistribusi semiparametrik karena dalam model Cox tidak memerlukan informasi tentang distribusi yang mendasari waktu survival dan untuk mengestimasi parameter regresi dari model Cox tanpa harus menentukan fungsi hazard dasar (Lee & Wang, 2003: 298).

Melalui model Cox dapat dilihat hubungan antara variabel bebas (variabel independen) terhadap variabel terikat (variabel dependen) yaitu waktu survival melalui fungsi hazardnya. Risiko kematian individu pada waktu tertentu

16

$$\log \frac{h(t, \mathbf{x})}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2.15)$$

Estimasi parameter regresi ($\beta_1, \beta_2, \dots, \beta_p$) pada model Cox dilakukan tanpa mengestimasi fungsi hazard dasar. Model pada persamaan (2.15) merupakan model dari *log hazard ratio (HR)*. *Hazard ratio* didefinisikan sebagai hazard dari satu individu dibagi dengan hazard individu yang berbeda (Kleinbaum & Klein, 2005: 24). Persamaan (2.15) dapat dinyatakan sebagai berikut,

$$\log[HR(\mathbf{x})] = (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p). \quad (2.16)$$

Persamaan (2.16) mengimplikasikan bahwa dalam model dengan variabel bebas x_1, x_2, \dots, x_p dan koefisien β_j yaitu peningkatan pada *log hazard ratio* untuk peningkatan satu satuan variabel bebas x_j , dengan asumsi bahwa nilai dari variabel bebas yang lain konstan. Dengan kata lain $\exp(\beta_j)$ adalah *hazard ratio* untuk peningkatan satu satuan dalam x_j . Variabel bebas dengan *hazard ratio* kurang dari 1 ($\beta < 0$), peningkatan nilai variabel bebas berhubungan dengan lebih menurunnya risiko dan lebih panjangnya waktu bertahan hidup. Ketika *hazard ratio* lebih besar dari 1 ($\beta > 0$), peningkatan nilai variabel bebas berhubungan dengan peningkatan risiko dan lebih pendeknya waktu bertahan hidup (Vittinghoff, Glidden, Shiboski, & McCulloch, 2004: 207).

1. Estimasi Parameter Model Cox Proportional Hazard

Parameter β_j pada model *Cox Proportional Hazard* akan diestimasi dengan menggunakan metode *Maximum Partial Likelihood Estimation (MPLE)*.

18

Diasumsikan hanya terdapat satu individu yang mengalami kematian pada tiap waktu kegagalan, jadi tidak terjadi *ties* pada data. *Ties* adalah keadaan dimana terdapat dua individu atau lebih yang mengalami kejadian gagal pada waktu yang sama. Hal lain yang perlu dipertimbangkan adalah peluang kematian suatu individu yang mati pada waktu kegagalan t_j , dengan syarat t_j menjadi salah satu yang diamati dari r waktu kegagalan t_1, t_2, \dots, t_r . Jika vektor variabel bebas dari individu yang mati pada waktu t_j dinotasikan dengan x_j , maka peluangnya menjadi sebagai berikut.

P [individu dengan variabel x_j mati pada t_j satu kematian pada t_j].

Seperti pada persamaan (2.8), misalkan kejadian A adalah individu dengan variabel x_j meninggal pada saat t_j dan kejadian B adalah semua kematian pada saat t_j , maka

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{P[\text{individu dengan variabel } x_j \text{ mati pada } t_j]}{P[\text{semua kematian pada } t_j]} \quad (2.17)$$

Pembilang pada persamaan (2.17) adalah bentuk sederhana dari risiko kematian pada waktu t_j untuk individu dengan variabel x_j . Jika pembilang

19

$$\begin{aligned} \ln L(\beta) &= \ln \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \\ &= \sum_{i=1}^r \left[\ln \left(\exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) - \ln \left(\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) \right] \\ &= \sum_{i=1}^r \left[\left(\sum_{j=1}^p \beta_j x_{ij} \right) - \ln \left(\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) \right]. \quad (2.20) \end{aligned}$$

Turunan pertama dari $\ln L(\beta)$ terhadap β_j yaitu sebagai berikut,

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta_j} &= \frac{\partial \left(\sum_{i=1}^r \left[\left(\sum_{j=1}^p \beta_j x_{ij} \right) - \ln \left(\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) \right] \right)}{\partial \beta_j} \\ \frac{\partial \ln L(\beta)}{\partial \beta_j} &= \sum_{i=1}^r \left[\sum_{j=1}^p x_{ij} - \frac{\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \sum_{j=1}^p x_{ij}}{\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right)} \right]. \quad (2.21) \end{aligned}$$

Pendugaan β_j dapat diperoleh dengan memaksimumkan turunan pertama fungsi *log partial likelihood* yaitu dengan mencari solusi dari:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = 0$$

$$\sum_{i=1}^r \left[\sum_{j=1}^p x_{ij} - \frac{\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \sum_{j=1}^p x_{ij}}{\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right)} \right] = 0. \quad (2.22)$$

Persamaan (2.22) diatas dapat diselesaikan secara numerik yaitu menggunakan metode Newton-Raphson.

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

21

tersebut adalah individu ke- i yang meninggal pada saat t_j , fungsi hazard ini dapat ditulis menjadi $h_i(t_j)$. Penyebut pecahan pada persamaan (2.17) adalah penjumlahan dari peluang kematian pada waktu t_j (dinotasikan $h_i(t_j)$) dari semua individu yang mempunyai risiko kematian pada waktu t_j . $R(t_j)$ adalah himpunan individu yang berisiko pada waktu t_j yang terdiri dari individu-individu yang bertahan hidup hingga t_j . Peluang dalam persamaan (2.17) menjadi $\frac{h_i(t_j)}{\sum_{i \in R(t_j)} h_i(t_j)}$, sehingga diperoleh

$$\begin{aligned} P(A|B) &= \frac{h_i(t_j)}{\sum_{i \in R(t_j)} h_i(t_j)} \\ P(A|B) &= \frac{h_0(t) \exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{i \in R(t_j)} h_0(t) \exp(\sum_{j=1}^p \beta_j x_{ij})} \\ &= \frac{h_0(t) \exp(\sum_{j=1}^p \beta_j x_{ij})}{h_0(t) \sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \\ P(A|B) &= \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})}. \quad (2.18) \end{aligned}$$

Berdasarkan hasil peluang bersyarat diatas, diperoleh fungsi *partial likelihood* sebagai berikut,

$$L(\beta) = \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})}. \quad (2.19)$$

Kemudian dari persamaan (2.19) diperoleh fungsi *log partial likelihood* yaitu sebagai berikut,

20

$$\begin{aligned} \frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} &= \frac{\partial}{\partial \beta_j} \left(\frac{\partial \ln L(\beta)}{\partial \beta_j} \right) \\ &= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^r \left[\sum_{j=1}^p x_{ij} - \frac{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \sum_{j=1}^p x_{ij}}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \right] \right] \\ &= \sum_{i=1}^r \left[\frac{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \sum_{j=1}^p x_{ij} \right)^2}{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \right)^2} - \frac{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \left(\sum_{j=1}^p x_{ij} \right)^2}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \right] \\ &= - \sum_{i=1}^r \left[\frac{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \left(\sum_{j=1}^p x_{ij} \right)^2}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} - \frac{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \right)^2}{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right]. \quad (2.23) \end{aligned}$$

Negatif turunan kedua dari *log likelihood* yaitu sebagai berikut,

$$\begin{aligned} - \frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} &= - \left[- \sum_{i=1}^r \left[\frac{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \left(\sum_{j=1}^p x_{ij} \right)^2}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} - \frac{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \right)^2}{\left(\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij}) \right)^2} \right] \right]. \quad (2.24) \end{aligned}$$

2. Prosedur Newton Raphson untuk Penaksiran Parameter Model Cox *Proportional Hazard*

Untuk memaksimumkan fungsi *partial likelihood* dalam penaksiran parameter model *Cox Proportional Hazard* dapat menggunakan prosedur Newton Raphson. Misalkan $L(\beta)$ merupakan fungsi *partial likelihood* p dimensi vektor $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$. Misalkan $U(\beta)$ merupakan vektor berukuran p dari turunan parsial pertama $L(\beta)$ seperti pada persamaan (2.25) berikut.

22

$$I(\beta) = (I_{ij}(\beta)), \quad i, j = 1, \dots, p \quad (2.26)$$

dengan memisalkan $I_{ij}(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial \beta_i \partial \beta_j}$.

$$I(\beta) = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{(\partial \beta_1)^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_p} \\ \frac{\partial^2 \ln L(\beta)}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 \ln L(\beta)}{(\partial \beta_2)^2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L(\beta)}{\partial \beta_p \partial \beta_1} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_p \partial \beta_2} & \dots & \frac{\partial^2 \ln L(\beta)}{(\partial \beta_p)^2} \end{bmatrix}$$

Algoritma metode Newton Raphson yaitu sebagai berikut.

$$\hat{\beta}_{c+1} = \hat{\beta}_c - I(\hat{\beta}_c)^{-1} U(\hat{\beta}_c) \quad (2.27)$$

dengan memisalkan, $c = 0, 1, 2, \dots$ dan $I^{-1}(\hat{\beta}_c)$ merupakan invers dari $I(\hat{\beta}_c)$.

Langkah iterasi dengan metode Newton Raphson sebagai berikut.

1. Menentukan nilai awal, $\hat{\beta}_0 = 0$.
2. $\hat{\beta}_1 = \hat{\beta}_0 - I(\hat{\beta}_0)^{-1} U(\hat{\beta}_0)$.
3. Iterasi dilakukan sampai memperoleh nilai yang konvergen, $\hat{\beta}_{c+1} \cong \hat{\beta}_c$.

ratio, uji *Wald*, dan uji *score*. Pengujian signifikansi parameter bertujuan untuk memeriksa apakah variabel bebas memiliki pengaruh nyata dalam model.

a. Uji *partial likelihood ratio*

Untuk menguji hipotesis bahwa satu atau beberapa parameter regresi β_j adalah nol dapat menggunakan uji *partial likelihood ratio* dinotasikan dengan G . Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p . Berikut langkah-langkah uji *partial likelihood* rasio:

1. Hipotesis:
 $H_0: \forall \beta_j = 0, j=1, \dots, p$
 $H_1: \exists \beta_j \neq 0, j=1, \dots, p$
2. Taraf signifikansi: α
3. Statistik uji:

$$G = -2[\ln L(0) - \ln L(\hat{\beta}_j)] \quad (2.31)$$

dengan,

$\ln L(0) = \log$ *partial likelihood* dari model tanpa variabel bebas (model *null*)
 $\ln L(\hat{\beta}_j) = \log$ *partial likelihood* dari model yang terdiri dari p variabel bebas.

4. Daerah penolakan:
 H_0 ditolak jika $G \geq \chi^2_{(\alpha; df=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.
5. Kesimpulan:
jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival (variabel dependen).

Variansi dari β_j dapat didefinisikan (Hosmer, Lemeshow, & May, 2008:

72) sebagai berikut,

$$Var(\hat{\beta}) = I(\hat{\beta})^{-1}, \quad (2.28)$$

Standar deviasi dari $\hat{\beta}_j$ merupakan akar kuadrat dari varians $\hat{\beta}_j$ (Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut,

$$SE(\hat{\beta}) = \sqrt{Var(\hat{\beta})} = \sqrt{I(\hat{\beta})^{-1}}, \quad (2.29)$$

Standar deviasi dapat digunakan untuk mencari selang kepercayaan $\hat{\beta}_j$ yaitu $(1 - \alpha)100\%$ selang kepercayaan untuk $\hat{\beta}_j$ (Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut,

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} SE(\hat{\beta}), \quad (2.30)$$

3. Pengujian Parameter

Melalui model Cox dapat diketahui hubungan antara variabel bebas (variabel independen) terhadap variabel terikat (variabel dependen) yaitu waktu survival melalui fungsi hazardnya, seperti yang ditunjukkan pada persamaan (2.13) yaitu,

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p),$$

Menurut David W. Hosmer dan Standley Lemeshow (2008: 89), terdapat tiga cara untuk menguji signifikansi parameter yaitu dengan uji *partial likelihood*

b. Uji *Wald*

Uji *Wald* digunakan untuk menguji pengaruh parameter secara terpisah, dinotasikan dengan z . Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p . Berikut langkah-langkah uji *Wald*:

1. Hipotesis:
 $H_0: \beta_j = 0$
 $H_1: \beta_j \neq 0$
2. Taraf signifikansi: α
3. Statistik uji:

$$z^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \quad (2.32)$$

4. Daerah penolakan:
 H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha; df=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.
5. Kesimpulan:
jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival (variabel dependen).
- c. Uji *Score*

Uji yang lain untuk menguji signifikansi parameter yaitu uji *Score*. Statistik uji ini adalah rasio dari turunan *log partial likelihood* pada persamaan (2.21), dengan akar kuadrat dari persamaan (2.26) semuanya dievaluasi terhadap $\beta_j = 0$.

$$H_1: \beta_j \neq 0$$

2. Taraf signifikansi: α

3. Statistik uji:

$$z^* = \left(\frac{\partial L \beta_j / \partial \beta_j}{\sqrt{I(\beta_j)}} \Big|_{\beta_j = 0} \right)^2 \quad (2.3.3)$$

4. Daerah penolakan:

H_0 ditolak jika $z^* \geq \chi^2_{(\alpha, df=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Kesimpulan:

jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival (variabel dependen).

4. Pemilihan Model Cox Terbaik

Pemilihan model terbaik diawali dengan pemilihan variabel yang masuk atau keluar dari model. Menurut David Collett (2003: 61), pemilihan variabel yang masuk atau keluar dari model dapat dilakukan dengan tiga cara yaitu seleksi *forward*, eliminasi *backward* dan prosedur *stepwise*. Prosedur seleksi *stepwise* merupakan kombinasi dari dua proses yaitu seleksi *forward* dan seleksi

27

$p^{(0)}(e_1) < \alpha$, α merupakan taraf signifikansi yang dipilih, maka proses berlanjut pada langkah 1.

Langkah 1: Langkah ini dimulai dengan variabel X_{e_1} dalam model. Hitung

$$G^{(1)}(i) = -2[\ln L^{(1)}(i) - \ln L(X_{e_1})] \text{ dengan } i = 1, 2, 3, \dots, p$$

dan $i \neq e_1$. $\ln L(X_{e_1})$ adalah *log partial likelihood* dengan dua variabel bebas dalam model, dan $\ln L^{(1)}(i)$ adalah *log partial likelihood* dengan variabel bebas yang terpilih pada langkah 0.

P-value untuk uji signifikansi dari penambahan X_i pada model yang terdiri dari variabel X_{e_1} yaitu $p^{(1)}(i) = P[\chi^2(1) \geq G^{(1)}(i)]$. Variabel bebas yang dipilih untuk masuk dalam langkah 2 yaitu X_{e_2} dengan $p^{(1)}(e_2) = \min_{i \neq e_1} p^{(1)}(i)$. Jika variabel yang terpilih yaitu X_{e_2} signifikan dengan $p^{(1)}(e_2) < \alpha$, maka berlanjut pada langkah 2.

Langkah 2: Langkah ini dimulai dengan variabel X_{e_1} dan X_{e_2} dalam model.

$$\text{Hitung } G^{(2)}(i) = -2[\ln L(X_{e_1}, X_{e_2}) - \ln L^{(2)}(i)] \text{ dengan}$$

$i = 1, 2, 3, \dots, p$ dan $i \neq e_1, e_2$, dengan $\ln L(X_{e_1}, X_{e_2})$ adalah

log partial likelihood dengan tiga variabel bebas dalam model,

$\ln L^{(2)}(i)$ adalah *log partial likelihood* dengan variabel bebas

yang terpilih pada langkah 1. *P-value* untuk uji signifikansi

dari penambahan variabel baru pada model yang terdiri dari

X_{e_1} dan X_{e_2} yaitu $p^{(2)}(i) = P[\chi^2(1) \geq G^{(2)}(i)]$. Variabel

bebas yang dipilih untuk masuk dalam langkah 3 yaitu X_{e_3}

29

backward. Seleksi *backward* atau seleksi mundur dengan memasukkan semua variabel kedalam model kemudian mengeluarkannya satu per satu jika variabel memiliki peningkatan nilai $-2\ln L(\beta)$ terbesar. Jika sudah tidak ada peningkatan nilai $-2\ln L(\beta)$ secara signifikan dari pengurangan variabel maka langkah *backward* dihentikan.

Dalam skripsi ini pemilihan model terbaik dilakukan menggunakan seleksi *forward*. Seleksi *forward* atau seleksi maju yaitu dengan menambahkan variabel satu demi satu dalam setiap langkahnya. Menurut David W. Hosmer dan Stanley Lemeshow (2008: 416) taraf signifikansi yang digunakan dalam seleksi *forward* disarankan antara 20%-25% untuk memungkinkan lebih banyak variabel yang masuk dalam model. Pada masing-masing tahapan, akan diputuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model. Berikut langkah-langkah dari seleksi *forward*:

Langkah 0: Misalkan ada sebanyak p variabel bebas X_i , dengan $i =$

$$1, 2, 3, \dots, p. \text{ Hitung } G^{(0)}(i) = -2[\ln L^{(0)}(0) - \ln L^{(0)}(i)]$$

dengan $\ln L^{(0)}(0)$ adalah *log partial likelihood* dari model nol

(model tanpa variabel bebas) dan $\ln L^{(0)}(i)$ adalah *log partial*

likelihood dari model dengan variabel bebas X_i , dengan *p-*

value untuk uji signifikansinya yaitu $p^{(0)}(i) = P[\chi^2(1) \geq$

$G^{(0)}(i)]$. Variabel bebas yang pertama kali masuk dalam

model adalah variabel bebas yang paling signifikan

berpengaruh terhadap waktu *survival* dimotaskan dengan X_{e_1}

yaitu variabel yang memiliki $p^{(0)}(e_1) = \min_i p^{(0)}(i)$ dengan

28

dengan $p^{(2)}(e_2) = \min_{i \neq e_1, e_2} p^{(2)}(i)$. Jika variabel yang

terpilih yaitu X_{e_2} signifikan dengan $p^{(2)}(e_2) < \alpha$, maka

berlanjut pada langkah 3.

Langkah 3: Langkah ini sama dengan langkah 2 dalam proses eliminasi

untuk menentukan variabel yang signifikan masuk ke dalam

model. Proses eliminasi variabel dilanjutkan seperti langkah 2

sampai langkah terakhir yaitu langkah S.

Langkah S: Pada langkah S, satu atau dua hal berikut terjadi yaitu: 1) semua

variabel sudah masuk dalam model dan tidak ada yang keluar,

2) setiap variabel bebas tidak masuk dalam model yaitu

mempunyai $p^{(S)}(i) < \alpha$. Pada langkah ini, tidak ada variabel

bebas yang terpilih untuk masuk dan tidak ada variabel bebas

yang keluar dari model.

Setelah diperoleh variabel yang masuk dalam model dengan beberapa langkah diatas, kemudian dilanjutkan dengan pemeriksaan apakah terdapat interaksi antar variabel tersebut dengan melakukan uji *likelihood ratio* yaitu membandingkan model Cox tanpa interaksi terhadap model Cox dengan penambahan variabel interaksi. Langkah-langkah pemilihan variabel interaksi yang masuk dalam model, dapat dilakukan dengan seleksi *forward*, seleksi *backward* maupun prosedur *stepwise* dengan langkah-langkah sama seperti yang telah dijelaskan dalam pemilihan variabel yang masuk kedalam model untuk mendapatkan model Cox terbaik.

30

waktu (Collett, 2003: 141). Pengujian asumsi proporsional hazard yang digunakan dalam skripsi ini yaitu dengan pendekatan grafik dengan plot *Log-Minus-Log Survival* dan dengan menggunakan residual *Schoenfeld*. Apabila ditemukan terdapat variabel yang tidak memenuhi asumsi proporsional hazard maka asumsi proporsional hazard dilanggar (*cox nonproportional hazard*). Menurut David Collett (2003: 182) ada 3 pilihan untuk mengatasi *cox nonproportional hazard* yaitu mengeluarkan variabel bebas yang tidak memenuhi asumsi dari model, menggunakan model *Cox Stratifikasi* dan dengan perluasan model Cox.

1. Pengujian Asumsi Proporsional Hazard dengan Grafik *Log-Minus-Log Survival*

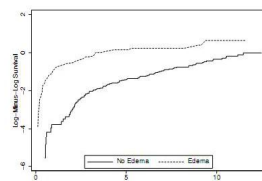
Pendekatan grafik yang digunakan yaitu dengan plot *Log-Minus-Log Survival* atau plot log-hazard kumulatif. Menurut model regresi Cox, fungsi hazard untuk kejadian setiap waktu t untuk individu $ke-i$ dapat dituliskan seperti pada persamaan (2.13) yaitu sebagai berikut.

$$h_i(t, x) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_j x_{ji})$$

$$= h_0(t) \exp\left(\sum_{j=1}^p \beta_j x_{ji}\right)$$

Survival adalah sebuah plot dari logaritma estimasi fungsi kumulatif hazard terhadap waktu survival, menghasilkan kurva paralel jika laju *proportional hazard* di seluruh kelompok berbeda. Kelemahan menggunakan pendekatan grafik/plot adalah bersifat subyektif, paralel atau tidaknya tergantung cara pandang peneliti (Collett, 2003: 142).

Sebagai contoh menurut Vittinghoff, Glidden, Shiboski, & McCulloch (2005: 235) misalnya pada kasus pengaruh edema pada penderita *primary biliary cirrhosis (PBC)*. *Primary biliary cirrhosis (PBC)* yaitu penyakit krusakan saluran-saluran kecil empedu di hati, yang menyebabkan empedu menumpuk di hati. Edema merupakan penyakit dimana seseorang mengalami peningkatan volume cairan pada kaki. Terdapat pasien penderita *primary biliary cirrhosis (PBC)* yang mengalami edema dan tidak mengalami edema. Berikut adalah gambar plot *Log-Minus-Log Survival* untuk variabel edema.



Gambar 2. 2 Plot *Log-Minus-Log Survival* Pada Variabel Edema

Pada Gambar 2.2 terlihat bahwa plot *Log-Minus-Log Survival* pada pasien dengan edema dan tidak mengalami edema mendekati *parallel*, sehingga

$$h_i(x_i) = h_0(t) \exp(\beta x_i), \quad (2.34)$$

x_i merupakan vektor dari nilai variabel bebas untuk individu tersebut, β merupakan vektor dari parameter, dan $h_0(t)$ merupakan fungsi hazard dasar. Apabila kedua sisi diintegrasikan dari nol hingga t , maka diperoleh sebagai berikut.

$$\int_0^t h_i(x_i) dx = \exp(\beta x_i) \int_0^t h_0(t) dx \quad (2.35)$$

dengan menggunakan persamaan (2.11), sehingga diperoleh

$$H_i(t) = \exp(\beta x_i) H_0(t) \quad (2.36)$$

dengan memisalkan $H_i(t)$ dan $H_0(t)$ merupakan fungsi kumulatif *hazard*. Selanjutnya dilakukan logaritma pada persamaan (2.36) untuk kedua sisi:

$$\log H_i(t) = \log \exp(\beta x_i) + \log H_0(t). \quad (2.37)$$

Menggunakan persamaan (2.12), persamaan (2.37) ekuivalen dengan persamaan berikut:

$$\log[-\log(S_i(t))] = \beta x_i + \log[-\log(S_0(t))]. \quad (2.38)$$

Persamaan (2.38) menunjukkan bahwa fungsi *Log-Minus-Log Survival* tidak bergantung terhadap waktu. Ini berarti bahwa fungsi *Log-Minus-Log Survival* pada model *Cox Proportional Hazard* pada persamaan (2.13) berlaku jika diplotkan melawan waktu survival dan kurva yang terbentuk akan paralel. Dalam menggunakan plot *Log-Minus-Log Survival* ini, data survival dikelompokkan sesuai dengan tingkat/ kategori faktor. Jika variabel kontinu maka nilainya perlu dikelompokkan menjadi variabel kategori. Plot *Log-Minus-Log*

32

mengindikasikan bahwa asumsi proporsional hazard pada variabel edema terpenuhi.

2. Pengujian Asumsi Proporsional Hazard dengan *Residual Schoenfeld*

Menurut David Collett (2003: 113) terdapat beberapa jenis residual yaitu residual *Cox-Snell*, residual *Martingale*, residual *Deviance*, residual *Schoenfeld* dan residual *Score*. Residual *Schoenfeld* akan dibahas dalam skripsi ini. Residual *Schoenfeld* untuk individu $ke-i$ pada variabel bebas $ke-j$ yaitu sebagai berikut.

$$\hat{r}_{ij} = c_j \left(x_{ij} - \frac{\sum_{l \in R(t_i)} x_{lj} \exp(\beta' x_{lj})}{\sum_{l \in R(t_i)} \exp(\beta' x_{lj})} \right). \quad (2.39)$$

Misalkan c_j indikator penyensoran, $c_j = 1$ untuk tidak tersensor, $c_j = 0$ untuk yang lainnya, x_{ij} adalah nilai dari variabel bebas $ke-j$, $j = 1, 2, \dots, p$ untuk individu $ke-i$ dalam pengamatan dan $R(t_i)$ merupakan himpunan semua individu yang berisiko saat t_i . Vektor dari p residual *Schoenfeld* untuk individu $ke-i$ yaitu sebagai berikut,

$$\hat{r}_i^* = (\hat{r}_{i1}, \hat{r}_{i2}, \dots, \hat{r}_{ip}). \quad (2.40)$$

Vektor dari *scaled residual Schoenfeld* yaitu sebagai berikut,

$$\hat{r}_i^* = r \text{ var}(\hat{\beta}) \hat{r}_i \quad (2.41)$$

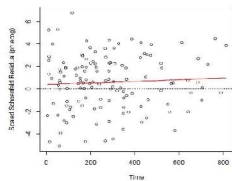
Misalkan r adalah banyaknya kejadian dari n individu dan $\text{var}(\hat{\beta})$ adalah ragam dari parameter β .

Cox Proportional Hazard dikatakan proporsional jika rasio dari hazard independen terhadap waktu. Jika terdapat variabel bebas yang tergantung pada

$$E(r_{pi}^*) \approx \beta_j(t_j) - \hat{\beta}_j \quad (2.42)$$

dengan memisalkan $\beta_j(t)$ adalah nilai dari parameter pada saat waktu kejadian ke- i . $\hat{\beta}_j$ merupakan nilai estimasi dari β_j pada model regresi cox. Plot dari nilai $r_{pi}^* + \hat{\beta}_j$ melawan waktu kejadian memberikan informasi tentang bentuk dari koefisien yang bergantung pada waktu dari X_j yaitu $\beta_j(t)$. Apabila plot tersebut horizontal maka mengindikasikan bahwa koefisien dari X_j konstan, dan asumsi proporsional hazard terpenuhi.

Sebagai contoh pada kasus penyakit kanker paru-paru. Berikut adalah gambar plot residual *Schoenfeld* pada variabel nilai *ecog* (Therneau & Grambsch, 2000: 137).



Gambar 2.3 Plot Residual *Schoenfeld* Pada Variabel Nilai *Ecog*

35

$$\begin{aligned} \log\left(\frac{h(t, X_1)}{h(t, X_0)}\right) &= \log(e^{(X_1 - X_0)\beta}) \\ \log\left(\frac{h(t, X_1)}{h(t, X_0)}\right) &= (X_1 - X_0)\beta. \end{aligned} \quad (2.44)$$

Untuk nilai kovariat X_j lainnya tetap dapat diinterpretasikan seperti berikut

$$\log\left(\frac{h(t, X_{j+1})}{h(t, X_j)}\right) = \beta_j \quad (2.45)$$

Persamaan (2.45) dapat disimpulkan bahwa setiap naiknya nilai β_j akan memperbesar nilai *log hazard ratio*, sehingga diperoleh

$$\begin{aligned} \log\left(\frac{h(t, X_{j+1})}{h(t, X_j)}\right) &= \beta_j \\ \frac{h(t, X_{j+1})}{h(t, X_j)} &= e^{\beta_j}, \quad \forall t > 0. \end{aligned} \quad (2.46)$$

Dengan demikian nilai $\exp(\beta_j)$ merupakan hazard rasio yang dapat dihubungkan dengan kenaikan nilai x_j .

Berdasarkan persamaan $h(t, X_j) \approx P(t < T < t + \Delta t | T \geq t, X)$, maka persamaan (2.46) dapat dituliskan sebagai berikut,

$$\frac{P(t < T < t + \Delta t | T \geq t, X_{j+1})}{P(t < T < t + \Delta t | T \geq t, X_j)} = e^{\beta_j}, \quad \forall t > 0. \quad (2.47)$$

Jadi $\exp(\beta_j)$ dapat juga diinterpretasikan sebagai rasio 2 peluang bersyarat dari gagalnya individu yang diketahui tersebut masih hidup pada saat t . Persamaan (2.46) ekuivalen dengan,

37

Diperoleh data pasien dari sebuah studi variabel prognostik tentang perkembangan kanker paru-paru dari *North Central Cancer Treatment Group (NCCTG)* yang terdiri dari 228. Nilai *ecog* pada penyakit kanker paru merupakan nilai dari hasil salah satu sistem penilaian mengenai kondisi pasien kanker yang dikemukakan oleh *Eastern Cooperative Oncology Group (ECOG)*. Pada gambar 2.3 tersebut terlihat bahwa residual *Schoenfeld* pada variabel nilai *ecog* memiliki kemiringan mendekati nol atau mendekati horizontal, sehingga mengindikasikan bahwa asumsi proporsional hazard untuk variabel nilai *ecog* terpenuhi.

E. Interpretasi Model *Cox Proportional Hazard*

Ketika model cox $h(t, X_j) = h_0 t \exp(\beta X_j)$ telah terbentuk, maka dilakukan interpretasi koefisien regresi. Diperlukan hazard rasio agar koefisien regresi dapat diinterpretasikan. Berdasarkan persamaan (2.16), untuk variabel bebas X_0 dan X_1 dari dua individu maka diperoleh,

$$\frac{h(t, X_1)}{h(t, X_0)} = \frac{h_0 t \exp(\beta X_1)}{h_0 t \exp(\beta X_0)} = \frac{\exp(\beta X_1)}{\exp(\beta X_0)} = e^{(X_1 - X_0)\beta}, \quad \forall t > 0. \quad (2.43)$$

Persamaan (2.43) menunjukkan besarnya rasio relatif dari individu dengan faktor risiko X_1 dibandingkan dengan faktor risiko X_0 dari individu lain (Lee & Wang, 2003: 299). Persamaan (2.43) dapat dituliskan sebagai berikut,

$$\frac{h(t, X_1)}{h(t, X_0)} = e^{(X_1 - X_0)\beta}$$

36

$$\begin{aligned} \frac{h(t, X_{j+1})}{h(t, X_j)} - \frac{h(t, X_j)}{h(t, X_j)} &= e^{\beta_j} - 1 \\ \frac{h(t, X_{j+1}) - h(t, X_j)}{h(t, X_j)} &= e^{\beta_j} - 1, \quad \forall t > 0. \end{aligned} \quad (2.48)$$

Oleh karena itu $e^{\beta_j} - 1$ dapat diinterpretasikan sebagai persentase perubahan nilai *hazard* baik naik atau turun dari setiap naiknya nilai x_j , dengan menganggap kovariat yang lain tetap.

Terdapat 3 macam ketentuan tentang bertambahnya atau berkurangnya nilai hazard yaitu sebagai berikut.

1. $\beta_j > 0$, maka kenaikan nilai x_j akan memperbesar nilai hazard atau semakin besar risiko seorang individu untuk meninggal atau gagal.
2. $\beta_j < 0$, maka kenaikan nilai x_j akan memperkecil nilai hazard atau semakin kecil risiko seorang individu untuk meninggal atau gagal.
3. $\beta_j = 0$, maka besar risiko seorang individu untuk hidup sama dengan besarnya risiko seorang individu untuk meninggal atau gagal.

F. Kejadian Berulang

Analisis survival yang dilakukan dengan menggunakan model *Cox Proportional Hazard* memperhatikan kejadian tunggal yang dialami subyek penelitian. Dalam kondisi tersebut setiap subyek diteliti mengalami kejadian yang diharapkan peneliti (*failure time*) hanya sekali. Menurut Hosmer dan Lemeshow (2008: 307), dalam kondisi lain, tidak menutup kemungkinan bahwa kejadian yang diinginkan peneliti mungkin terjadi lebih dari satu kali dalam suatu subyek.

38

penelitian. Banyaknya interval waktu yang dimiliki subyek sesuai dengan banyaknya ulangan kejadian yang dialami selama penelitian. Perbedaan baris dalam susunan data kejadian berulang disebabkan oleh tiap-tiap subyek yang diasumsikan independen pada intervalnya, walaupun beberapa baris berasal dari subyek yang sama namun pengulangan yang terjadi dianggap berasal dari subyek yang berbeda.

Pada analisis survival kejadian yang berulang, subyek penelitian dikeluarkan dari himpunan risiko ketika subyek tersebut mengalami kejadian atau tersensor dan tetap berada pada himpunan risiko hingga akhir waktu interval waktu survivalnya. Model *Counting Process* mengkaji hubungan antara waktu survival dengan variabel-variabel yang diduga mempengaruhi waktu survival (variabel respon) akan memperhatikan perulangan kejadian yang dialami setiap subyek.

Kejadian berulang dalam analisis survival terdiri dari dua macam menurut Kleinbaum dan Klein (2005: 335), yaitu kejadian berulang identik dan kejadian berulang tidak identik. Kejadian berulang dikatakan identik jika urutan kejadian berulang tidak menimbulkan efek perbedaan tertentu. Sebagai contoh kasus serangan jantung, baik kasus serangan jantung yang pertama, kedua, dan selanjutnya dianggap sama, tidak menyebabkan tingkat keparahan yang berbeda,

Variabel bebas yang memenuhi asumsi proporsional hazard masuk ke dalam model, sedangkan variabel bebas yang tidak memenuhi asumsi, yang sedang distratifikasi, tidak masuk dalam model (Kleinbaum dan Klein, 2005: 181).

1. Stratifikasi Variabel dalam Model Cox Stratifikasi

Dalam model Cox Stratifikasi diasumsikan terdapat sebanyak p variabel bebas. Sebanyak k variabel bebas diantaranya memenuhi asumsi proporsional hazard dinotasikan X_1, X_2, \dots, X_k dengan $k < p$. Variabel bebas yang tidak memenuhi asumsi proporsional hazard sebanyak m yang diperoleh dari $p - k = m$ yaitu $X_{k+1}, X_{k+2}, \dots, X_p$ yang dinotasikan Z_1, Z_2, \dots, Z_m .

$$X_{k+1} \longrightarrow Z_1, X_{k+2} \longrightarrow Z_2, \dots, X_p \longrightarrow Z_m$$

Variabel yang tidak memenuhi asumsi proporsional hazard Z_i dengan $i = 1, \dots, m$ dikeluarkan dari model Cox untuk dilakukan stratifikasi terhadap variabel tersebut sehingga diperoleh variabel stratifikasi Z^* . Variabel bebas yang memenuhi asumsi proporsional hazard akan masuk ke dalam model Cox Stratifikasi. Meskipun begitu variabel bebas yang dikeluarkan dari model tetap memiliki peran dan dengan dilakukan stratifikasi variabel akan terlihat kontribusi masing-masing variabel bebas tersebut dalam strata yang berbeda.

Langkah awal yang harus dilakukan adalah identifikasi variabel bebas yang tidak memenuhi asumsi proporsional hazard. Terdapat k variabel bebas yang memenuhi asumsi proporsional hazard, dinotasikan X_1, X_2, \dots, X_k , diantara p variabel bebas yang telah diasumsikan, yang diantaranya terdapat sebanyak $p - k = m$ variabel bebas yang tidak memenuhi asumsi proporsional hazard. Langkah selanjutnya adalah melakukan stratifikasi variabel. Dalam stratifikasi

tidak dibedakan tipenya maka dianggap kejadian berulang identik. Pada kejadian berulang identik ini analisis yang digunakan adalah model *Counting Process* yang dikembangkan oleh Anderson Gill (Andersen et al. 1993).

Kejadian berulang dikatakan tidak identik jika ada urutan kejadian berulang atau perbedaan kategori yang lain yang menyebabkan efek perbedaan tertentu pada kejadian berulang. Misal pada kasus penyakit kanker otak, apabila diasumsikan ada perbedaan urutan kejadian tumor, tumor pertama, kedua dan selanjutnya, maka kejadian tumor berulang tersebut tidak identik, dapat dikategorikan. Pada kejadian berulang tidak identik ini analisis survival menggunakan pendekatan model Cox Stratifikasi salah satunya model Cox Stratifikasi *PWP-Gap time* yang dikemukakan oleh Prentice, William dan Peterson.

G. Model Cox Stratifikasi

Model Cox Stratifikasi merupakan perluasan dari model *Cox Proportional Hazard* untuk mengatasi variabel bebas yang tidak memenuhi asumsi proporsional hazard. Asumsi proporsional hazard menyatakan bahwa rasio fungsi hazard dari dua individu konstan dari waktu ke waktu atau ekuivalen dengan pernyataan bahwa fungsi hazard suatu individu terhadap fungsi hazard individu lain adalah proporsional (Guo, 2010: 77). Jika suatu variabel bergantung pada terhadap waktu, yaitu variabel bebas yang diamati lebih dari sekali selama penelitian dan nilainya berubah dari waktu ke waktu, maka asumsi proporsional hazard dilanggar (Allison, 2010: 179). Modifikasi dilakukan dengan menstratifikasi variabel bebas yang tidak memenuhi asumsi proporsional hazard.

variabel langkah pertama yang dilakukan adalah mendefinisikan variabel baru yang akan digunakan untuk stratifikasi yaitu Z^* dengan membentuk kategori dari masing-masing Z_i . Variabel stratifikasi Z^* akan memiliki m^* kategori yaitu banyaknya kombinasi (strata) yang terbentuk setelah mengkategorikan masing-masing Z_i .

Cara pendefinisian variabel baru yang disebut dengan Z^* adalah sebagai berikut (Kleinbaum dan Klein, 2005: 180):

- Membuat kategori untuk masing-masing variabel yang tidak memenuhi asumsi proporsional hazard Z_i , dengan $i = 1, 2, \dots, m$.
- Membentuk kombinasi dengan kategori-kategori tersebut.
- Membentuk strata sebanyak kombinasi yang terbentuk.

Berdasarkan skala pengukurannya, variabel terbagi atas dua yaitu variabel kategorik dan variabel numerik. Stratifikasi variabel Z_i dapat dilakukan dengan ketentuan sebagai berikut:

- Z_i merupakan variabel kategorik
Apabila Z_i berupa variabel kategorik maka tidak perlu dilakukan kategorisasi karena banyaknya kategori sudah diketahui. Variabel kategorik atau dikenal dengan istilah kualitatif merupakan variabel yang memuat skala pengukuran nominal dan ordinal.
- Z_i merupakan variabel numerik
Apabila Z_i berupa variabel numerik, maka perlu dilakukan kategorisasi. Dalam kategorisasi atau pengkategorian, tidak terdapat ketentuan khusus untuk menentukan banyaknya kategori. Banyaknya kategori biasanya disesuaikan

asumsi proporsional hazard yaitu variabel umur (Z_1) dan variabel tinggi badan (Z_2). Variabel umur terdiri dari 3 kategori yaitu anak-anak, dewasa dan tua, sedangkan variabel tinggi badan terdiri dari 2 kategori yaitu pendek dan tinggi. Selanjutnya terbentuk kategori baru hasil kombinasi dari 3 kategori dari variabel umur dan 2 kategori dari variabel tinggi badan. Terdapat 6 kategori baru yang dimiliki variabel baru Z^* . Secara umum, apabila terdapat variabel bebas Z_1, Z_2, \dots, Z_m yang masing-masing variabel tersebut memiliki kategori sebanyak k_1, k_2, \dots, k_m , sehingga variabel baru Z^* memiliki kategori sebanyak $m^* = k_1 \times k_2 \times \dots \times k_m$.

2. Pembentukan model Cox Stratifikasi

Menurut Klain dan Kleinbaum (2005: 181) bentuk umum fungsi hazard dari model cox stratifikasi adalah sebagai berikut:

$$h_s(t, X) = h_{0s}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k] \quad (2.49)$$

dengan,

$$\begin{aligned} s &= \text{strata yang didefinisikan dari } Z^*, s=1, 2, \dots, m^* \\ h_{0s}(t) &= \text{fungsi dasar hazard untuk setiap strata} \\ \beta_1, \beta_2, \dots, \beta_k &= \text{parameter regresi} \end{aligned}$$

Strata didefinisikan sebagai kategori yang berbeda dari variabel stratifikasi Z^* dan m^* merupakan banyaknya strata. Dalam model Cox Stratifikasi, fungsi dasar

dengan masing-masing variabel bebas pada model persamaan (2.5) dapat juga dituliskan sebagai berikut:

$$\begin{aligned} h_s(t, X) &= h_{0s}(t) \exp[\beta_1 X_1 + \dots + \beta_k X_k + \beta_{k+1}(Z_1^* \times X_1) + \dots + \beta_{k+2}(Z_1^* \times X_k) \\ &+ \beta_{k+3}(Z_2^* \times X_1) + \dots + \beta_{k+4}(Z_2^* \times X_k) + \beta_{k+m^*-1}(Z_{m^*-1}^* \times X_1) + \dots \\ &+ \beta_{k+m^*-1}(Z_{m^*-1}^* \times X_k)] \end{aligned} \quad (2.51)$$

dengan $s = 1, 2, \dots, m^*$, strata yang terdefinisi dari Z^* . Untuk menuliskan model alternatif tersebut dengan benar, didefinisikan $m^* - 1$ variabel *dummy* untuk membedakan m^* kategori dari Z^* . Variabel *dummy* yang disebut juga variabel boneka atau variabel kategori, pada model diatas terdapat sebanyak $m^* - 1$ karena strata atau kategori sebanyak m^* . Variabel *dummy* dinotasikan $Z_1^*, Z_2^*, \dots, Z_{m^*-1}^*$ merupakan variabel yang terlibat dengan perkalian masing-masing X , bernilai 1 untuk variabel yang diteliti dan bernilai 0 untuk yang lain.

3. Metode Penaksiran Parameter Model Cox Stratifikasi

Estimasi parameter pada model cox stratifikasi ini menggunakan metode seperti halnya metode *Maximum Partial Likelihood Estimation (MPLE)* pada model *cox proportional hazard*, yang disebut dengan *Maximum Stratified Partial Likelihood Estimation*. Pendugaan parameter regresi dengan metode *MPLE* adalah nilai ketika fungsi *partial likelihood* maksimum. Dalam model cox stratifikasi akan ditentukan peluang kematian suatu individu ke i yang mati pada strata ke s pada waktu kegagalan t_{si} . Jika k variabel bebas individu yang mati pada waktu t_{si} , dinotasikan dengan $x_{(s)i}$ maka peluangnya menjadi sebagai berikut:

$$P[\text{individu dengan variabel } x_{(s)i} \text{ mati pada saat } t_{si} | \text{satu kematian pada } t_{si}]$$

hazard $h_{0s}(t)$ berbeda untuk setiap strata. Terdapat model interaksi dan model tanpa interaksi yang terdefinisi dalam konsep model Cox Stratifikasi sebagai berikut:

a. Model tanpa Interaksi

Model Cox Stratifikasi tanpa interaksi ini merupakan bentuk umum dari model Cox Stratifikasi yang menunjukkan bahwa tidak ada interaksi antara variabel bebas seperti pada persamaan (2.49) yaitu sebagai berikut:

$$h_s(t, X) = h_{0s}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k]$$

Parameter regresi $\beta_1, \beta_2, \dots, \beta_k$ untuk model ini sama untuk setiap strata sehingga perkiraan rasio hazard sama untuk masing-masing strata. Untuk memperoleh perkiraan parameter regresi $\beta_1, \beta_2, \dots, \beta_k$ dengan memaksimalkan fungsi *partial likelihood* stratifikasi yang diperoleh dengan mengalikan bersama-sama fungsi *partial likelihood* dari masing-masing strata. Apabila dalam suatu kasus asumsi model tanpa interaksi ini tidak terpenuhi maka digunakan model interaksi Cox Stratifikasi.

b. Model dengan Interaksi

Model Cox Stratifikasi dengan interaksi antara variabel Z^* dengan X dalam model yang ditunjukkan sebagai berikut:

$$h_s(t, X) = h_{0s}(t) \exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{ks} X_k] \quad (2.50)$$

dengan $s = 1, 2, \dots, m^*$, strata yang terdefinisi dari Z^* .

Subscript s pada parameter regresi persamaan (2.50) menunjukkan bahwa parameter regresi berbeda-beda untuk masing-masing strata. Cara alternatif untuk menuliskan model interaksi dengan menggunakan perkalian yang melibatkan Z^*

Misalkan kejadian A adalah individu dengan variabel $x_{(s)}$ mati pada saat t_{si} dan kejadian B adalah semua kematian pada t_{si} maka

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} \\ &= \frac{P[\text{individu dengan variabel } x_{(s)} \text{ mati saat } t_{si}]}{P[\text{semua kematian saat } t_{si}]} \end{aligned} \quad (2.52)$$

Pada persamaan (2.52) diatas, pembilang merupakan bentuk sederhana dari risiko kematian untuk individu dengan variabel $x_{(s)}$ saat t_{si} yang merupakan fungsi hazard yang dinotasikan $h(t_{si})$. Penyebut pecahan pada persamaan (2.52) merupakan jumlahan peluang kematian atau fungsi hazard pada waktu t_{si} dari semua individu yang mempunyai risiko kematian pada waktu t_{si} . $R(t_{si})$ adalah himpunan individu yang berisiko pada waktu t_{si} yang terdiri dari individu-individu yang bertahan hidup hingga t_{si} . Peluang dalam persamaan (2.52) pada strata s menjadi

$$\begin{aligned} P(A|B) &= \frac{h(t_{si})}{\sum_{j \in R(t_{si})} h(t_{si})} \\ P(A|B) &= \frac{h_{0s}(t_{si}) \exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{si})} h_{0s}(t_{si}) \exp[\beta x_{(s)j}]} \\ &= \frac{h_{0s}(t_{si}) \exp[\beta x_{(s)i}]}{h_{0s}(t_{si}) \sum_{j \in R(t_{si})} \exp[\beta x_{(s)j}]} \\ &= \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{si})} \exp[\beta x_{(s)j}]} \end{aligned} \quad (2.53)$$

Dengan menggunakan hasil peluang bersyarat pada persamaan (2.53) maka diperoleh fungsi *partial likelihood* untuk setiap strata (*subscript s* yang mengindikasikan strata) sebagai berikut:

masing-masing fungsi *partial likelihood* dari setiap strata berasal dari fungsi hazard yang sesuai.

$$L_p(\beta) = \prod_{s=1}^{m^*} L_s(\beta) \quad (2.55)$$

$$= L_1(\beta) \times L_2(\beta) \times \dots \times L_{m^*}(\beta). \quad (2.56)$$

Kemudian diperoleh bentuk fungsi *log partial likelihood* stratifikasi sebagai berikut:

$$\begin{aligned} \ln L_p(\beta) &= \ln L_1(\beta) \times \ln L_2(\beta) \times \dots \times \ln L_{m^*}(\beta) \\ &= \ln \left(\prod_{s=1}^{m^*} L_s(\beta) \right) \\ \ln L_p(\beta) &= \ln \left[\prod_{s=1}^{m^*} \left(\prod_{i=1}^{n_s} \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}]} \right) \right] \\ &= \sum_{s=1}^{m^*} \left[\ln \left(\prod_{i=1}^{n_s} \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}]} \right) \right] \\ \ln L_p(\beta) &= \sum_{s=1}^{m^*} \left[\sum_{i=1}^{n_s} \beta x_{(s)i} - \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) \right]. \quad (2.57) \end{aligned}$$

Untuk mendapatkan estimasi parameter regresi $\beta = \beta_1, \beta_2, \dots, \beta_k$ dengan memaksimalkan fungsi *partial likelihood* yaitu dengan menyelesaikan turunan

47

Turunan kedua persamaan dari fungsi *log partial likelihood* Model Cox Stratifikasi pada persamaan (2.57) adalah sebagai berikut :

$$\begin{aligned} \frac{\partial^2}{\partial \beta^2} \ln L_p(\beta) &= \frac{\partial}{\partial \beta} \left(\frac{\partial}{\partial \beta} \ln L_p(\beta) \right) \\ \frac{\partial^2}{\partial \beta^2} \ln L_p(\beta) &= \frac{\partial}{\partial \beta} \left[\frac{\partial}{\partial \beta} \sum_{s=1}^{m^*} \sum_{i=1}^{n_s} \beta x_{(s)i} - \frac{\partial}{\partial \beta} \sum_{s=1}^{m^*} \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) \right] \quad (2.60) \end{aligned}$$

4. Prosedur Newton Raphson untuk Penaksiran Parameter Model Cox Stratifikasi

Misalkan $L_p(\beta)$ merupakan fungsi *partial likelihood* k dimensi dengan vektor $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$. Misalkan $U(\beta)$ merupakan vektor ukuran p diperoleh dari turunan parsial pertama $\ln L_p(\beta)$.

$$U(\beta) = [U_1(\beta), U_2(\beta), \dots, U_k(\beta)]^T \quad (2.61)$$

dimana $U_j(\beta) = \frac{\partial \ln L_p(\beta)}{\partial \beta_j}$, $j = 1, 2, \dots, k$.

Misalkan $I(\beta)$ adalah matriks Hessian berukuran $k \times k$ dari turunan partial kedua $\ln L_p(\beta)$ yaitu

$$I(\beta) = (I_{ij}(\beta)) \text{ dengan } i, j = 1, \dots, k \quad (2.62)$$

dengan $I_{ij}(\beta) = \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_i \partial \beta_j}$, sehingga diperoleh matriks Hessian sebagai berikut

$$I(\beta) = \begin{bmatrix} \frac{\partial^2 \ln L_p(\beta)}{(\partial \beta_1)^2} & \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_1 \partial \beta_2} & \dots & \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_1 \partial \beta_k} \\ \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 \ln L_p(\beta)}{(\partial \beta_2)^2} & \dots & \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_2 \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_k \partial \beta_1} & \frac{\partial^2 \ln L_p(\beta)}{\partial \beta_k \partial \beta_2} & \dots & \frac{\partial^2 \ln L_p(\beta)}{(\partial \beta_k)^2} \end{bmatrix}$$

Algoritma pada metode Newton Raphson adalah sebagai berikut

49

logaritma fungsi *partial likelihood* terhadap β_g sama dengan nol seperti pada persamaan (2.58) berikut:

$$\begin{aligned} \frac{\partial}{\partial \beta_g} \ln L_p(\beta) &= 0 \\ \frac{\partial}{\partial \beta_g} \sum_{s=1}^{m^*} \sum_{i=1}^{n_s} \beta x_{(s)i} - \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) &= 0 \quad (2.58) \end{aligned}$$

dengan $g = 1, 2, \dots, k$.

Estimasi parameter pada model cox stratifikasi dengan metode *Maximum Partial Likelihood Estimation (MPLE)* dengan mencari solusi dari:

$$\begin{aligned} \text{a. } \frac{\partial}{\partial \beta_1} \ln L_p(\beta) &= \frac{\partial}{\partial \beta_1} \ln \left(\prod_{i=1}^{n_1} \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}]} \right) = 0 \\ &\Leftrightarrow \frac{\partial}{\partial \beta_1} \left[\sum_{i=1}^{n_1} \beta x_{(s)i} - \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) \right] = 0 \\ \text{b. } \frac{\partial}{\partial \beta_2} \ln L_p(\beta) &= \frac{\partial}{\partial \beta_2} \ln \left(\prod_{i=1}^{n_2} \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}]} \right) = 0 \\ &\Leftrightarrow \frac{\partial}{\partial \beta_2} \left[\sum_{i=1}^{n_2} \beta x_{(s)i} - \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) \right] = 0 \end{aligned}$$

Secara umum diperoleh,

$$\begin{aligned} \frac{\partial}{\partial \beta_g} \ln L_p(\beta) &= \frac{\partial}{\partial \beta_g} \ln \left(\prod_{i=1}^{n_g} \frac{\exp[\beta x_{(s)i}]}{\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}]} \right) = 0 \\ &\Leftrightarrow \frac{\partial}{\partial \beta_g} \left[\sum_{i=1}^{n_g} \beta x_{(s)i} - \ln \left(\sum_{j \in R(t_{s,i})} \exp[\beta x_{(s)j}] \right) \right] = 0. \quad (2.59) \end{aligned}$$

Persamaan (2.59) diatas dapat diselesaikan secara numerik menggunakan metode Newton Raphson.

48

$$\hat{\beta}_{c+1} = \hat{\beta}_c - I(\hat{\beta}_c)^{-1} U(\hat{\beta}_c) \quad (2.63)$$

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

Langkah iterasi pada metode Newton Raphson sebagai berikut

- Menetapkan nilai awal: $\hat{\beta}_0 = 0$.
- Menghitung $\hat{\beta}_1 = \hat{\beta}_0 - I(\hat{\beta}_0)^{-1} U(\hat{\beta}_0)$.
- Iterasi dilakukan hingga memperoleh nilai yang konvergen: $\hat{\beta}_{c+1} \cong \hat{\beta}_c$.

Perkiraan varians dari perkiraan maksimum *partial likelihood* $\hat{\beta}$ merupakan invers dari persamaan (2.62) yaitu (Hosmer Lemeshow, 1999: 72) :

$$V\hat{a}r(\hat{\beta}) = I(\hat{\beta})^{-1}. \quad (2.64)$$

Untuk perkiraan standar deviasi dari $\hat{\beta}$ sebagai berikut,

$$SE(\hat{\beta}) = \sqrt{V\hat{a}r(\hat{\beta})}. \quad (2.65)$$

5. Asumsi Model Cox Stratifikasi dengan Interaksi

Untuk menguji asumsi model tanpa interaksi pada model Cox Stratifikasi digunakan uji *log partial likelihood ratio* yaitu dengan membandingkan *log partial likelihood* dalam model interaksi dan model tanpa interaksi. Berikut langkah-langkah uji *likelihood ratio*:

- Hipotesis dalam uji *log partial likelihood ratio* yaitu sebagai berikut:

$$H_0: \begin{cases} \beta_{11} = \dots = \beta_{k1} = 0 \\ \beta_{12} = \dots = \beta_{k2} = 0 \\ \vdots \\ \beta_{1m^*-1} = \dots = \beta_{km^*-1} = 0 \end{cases}$$

50

- Institute Inc.
- Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. 1993. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Asri, L.M.A. 2014. Penerapan Regresi Stratified Cox Dengan Metode Conditional 1 Pada Data Kejadian Berulang Tidak Identik. *Jurnal Matematika FMIPA Universitas Brawijaya Malang*.
- Collett, D. 2003. *Modelling Data in Medical Research, second edition*. London: Chapman and Hall.
- Fauzi Rahmawati. 2013. Model Cox Stratifikasi Untuk Mengatasi Nonproportional Hazard pada Analisis Survival. *Skripsi*. Universitas Negeri Yogyakarta.
- Feriana, D.A. 2011. Model Cox Stratifikasi. *Skripsi*. Universitas Indonesia
- Gudono. 2011. *Analisis Data Multivariat*. Yogyakarta: BPFE.
- Guo, S. 2010. *Survival Analysis*. New York: Oxford University Press, Inc.
- Hosmer D.W., dan Lemeshow S. 2008. *Applied Survival Analysis, Regression modelling of Time to Event Data*. Willey: New Jersey.
- Kalbfleisch, J.D., & Prentice, R.L. 2002. *The statistical Analysis of Failure Time Data, second edition*. New Jersey: John Wiley & Sons, Inc.
- Kelly, Patrick J. & Lim, Lynnete L-Y. 2000. Survival Analysis for Recurrent Event Data: An Application to Childhood Infectious Diseases. *Journal of Statistic in Medicine* Vol. 19.
- Kleinbaum, D. G., & Klein, M. 2005. *Survival Analysis A SelfLearning Text second edition*. USA: Springer Science+Business Media, Inc.
- Lawless, J. F. 2007. *Statistical Models and Method for Lifetime Data*. New York: John Wiley & Sons, Inc.

- Lee, E. T., & Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis Third Edition*. New Jersey: John Wiley & Sons, Inc.
- Lim, Hyun Ja, et.al. 2007. Comparison of Methods for Analyzing Recurrent Events Data: Application to the Emergency Department Visits of Pediatric Firearm Victims. *Journal of Accident Analysis and Prevention*. Vol. 39. 290-299.
- Liu, Xian. 2012. *Survival Analysis: Models and Applications*. New Jersey: John Wiley & Sons, Inc.
- Pinzon, R. & Asanti, L. 2010. *Awas Stroke! Pengertian, Gejala, Tindakan, Perawatan, & Pencegahan*. Yogyakarta: Andi
- Prentice, R.L., Williams, B.J., Peterson, A.V. 1981. On The Regression Analysis of Multivariate Failure Time Data. *Journal Biometrika*. Vol 68. 373-379
- Syahwan Udin. 2010. Regresi Cox. *Skripsi*. Yogyakarta: Universitas Negeri Yogyakarta.
- Therneau, T. M., & Grambsch, P. M. 2000. *Modeling Survival Data Extending The Cox Model*. New York: Springer_Verlag.
- Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. 2004. *Regression Methods in Biostatistics First Edition*. New York: Springer Science + Business Media.
- Yayasan Stroke Indonesia. 2011. *Sekilas Tentang Stroke*. Diakses tanggal 17 Januari 2015 dari <http://www.yastroki.or.id/>
- Yuliaji Siswanto. 2005. Beberapa Faktor Risiko yang Mempengaruhi Kejadian Stroke Berulang. *Tesis*. Semarang: Universitas Diponegoro Semarang