BAB III

PERLUASAN MODEL REGRESI COX PROPORTIONAL HAZARD DENGAN VARIABEL TERIKAT OLEH WAKTU

3.1 Model Regresi Cox Proportional Hazard dengan Variabel Terikat oleh Waktu

Model regresi *Cox proportional hazard* secara umum memuat satu atau lebih variabel prediktor bebas oleh waktu yang digunakan untuk memprediksi status dari variabel responnya. Sehingga model regresi *Cox proportional hazard* dapat dituliskan sebagai berikut:

$$h(t, X) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \beta_i X_i\right] \qquad \text{... (3.1.1)}$$

Model perluasan regresi *Cox proportional hazard* merupakan model yang melibatkan variabel prediktor bebas oleh waktu dan variabel prediktor terikat oleh waktu. Seperti halnya model regresi *Cox proportional hazard*, perluasan model ini memuat fungsi *baseline hazard* ($h_0(t)$) dikalikan dengan fungsi eksponensial. Dalam model perluasan regresi *Cox proportional hazard*, fungsi eksponensial memuat variabel prediktor bebas oleh waktu yang dinotasikan dengan X_i dan variabel prediktor terikat oleh waktu dinotasikan dengan $X_i g_i(t)$. Model perluasan regresi *Cox proportional hazard* ditunjukkan sebagai berikut:

$$h(t, X(t)) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \beta_i X_i + \sum_{i=1}^{p_2} \delta_i X_i g_i(t)\right] \qquad \text{... (3.1.2)}$$

Atau secara umum, model perluasan regresi *Cox proportional hazard* hanya memuat satu atau lebih variabel prediktor terikat oleh waktu yang dinotasikan dengan $X_i g_i(t)$ sehingga model perluasannya dapat dituliskan sebagai berikut:

$$h\big(t, X(t)\big) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \delta_i X_i g_i(t)\right] \qquad \ldots (3.1.3)$$

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

X_i = variabel prediktor bebas oleh waktu pada saat ke-i.

$X_i(t)$ = kumpulan dari variabel prediktor terikat oleh waktu pada saat ke-i. $\big(X_1, X_2, \ldots, X_{p_1},\ X_1(t), X_2(t), \ldots, X_{p_2}(t)\big)$.

$g_i(t)$ = fungsi waktu untuk variabel prediktor terikat oleh waktu ke-i.

dengan i = 1, 2, …

3.2 Rasio Hazard untuk Model Perluasan Regresi Cox Proportional Hazard

Bagian utama dari model ini adalah asumsi *proportional hazard* yang tidak terpenuhi ketika dalam penggunaan model perluasan regresi *Cox proportional hazard* dengan penambahan variabel terikat oleh waktu. Rasio *hazard* (HR) dalam model perluasan ini dideskripsikan sebagai rasio laju kegagalan pada saat t dari spesifikasi dua kelompok variabel prediktor. Dinotasikan dua kelompok variabel prediktor yaitu $X^*(t)$ dan $X(t)$ sehingga rasio *hazard* untuk model perluasan regresi *Cox proportional hazard* adalah sebagai berikut:

$$\widehat{\Psi}(t) = \frac{h_{X^*}(t)}{h_X(t)}$$

$$= \frac{h_0(t)\exp\left[\sum_{i=1}^{p_1} \beta_i X_i{}^* + \sum_{i=1}^{p_2} \delta_i X_i{}^* g_i(t)\right]}{h_0(t)\exp\left[\sum_{i=1}^{p_1} \beta_i X_i + \sum_{i=1}^{p_2} \delta_i X_i g_i(t)\right]}$$

$$= \exp\left[\sum_{i=1}^{p_1} \hat{\beta}_i [X_i^* - X_j] + \sum_{i=1}^{p_2} \hat{\delta}_i [X_i^*(t) - X_i(t)]\right] \qquad \text{... (3.2.1)}$$

Pada umumnya, rumusan rasio *hazard* melibatkan perbedaan nilai pada variabel terikat oleh waktu pada saat t sehingga rasio *hazard* memuat fungsi waktu. Dalam rasio *hazard* pada model perluasan regresi *Cox proportional hazard*, nilai koefisien $\hat{\delta}_i$ akan berbeda-beda pada setiap variabel terikat oleh waktu dan koefisien δ_i direpresentasikan sebagai pengaruh keseluruhan dari hubungan antara variabel-variabel terikat oleh waktu dengan mempertimbangkan semua waktu pada variabel yang digunakan dalam pengamatan. Contoh model perluasan regresi *Cox proportional hazard* dengan variabel terikat oleh waktu adalah sebagai berikut:

$$h\big(t, X(t)\big) = h_0(t) \exp[\beta X + \delta(X \times g(t))]$$

dengan $X = \begin{cases} 1 & metode\ A \\ 0 & metode\ B \end{cases}$

maka rasio hazardnya adalah

$$\hat{\Psi}(t) = \frac{h(t, X=1)}{h(t, X=0)}$$

$$= \frac{h_0(t)\exp\left[(1 \times \beta) + \delta(1 \times g(t))\right]}{h_0(t)\exp\left[(0 \times \beta) + \delta(0 \times g(t))\right]}$$

$$= \exp\left[\hat{\beta}(1-0) + \hat{\delta}\big((1 \times g(t)) - (0 \times g(t))\big)\right]$$

$$= \exp\left[\hat{\beta} + \hat{\delta} g(t)\right]$$

Dalam rumusan rasio *hazard*, apabila nilai $\widehat{\delta}_i > 0$ maka rasio *hazard* akan naik bersamaan dengan naiknya waktu yang mengakibatkan rasio *hazard* tidak konstan dan asumsi *proportional hazard* tidak terpenuhi untuk model perluasan ini.

3.3 Pendugaan Parameter

Penentuan taksiran regresi dan pengujian koefisien β dalam model perluasan regresi *Cox proportional hazard* dapat diestimasi dengan menggunakan prosedur maksimum likelihood. Asumsi dasar dalam pembentukan fungsi likelihood adalah waktu hidup dan waktu penyensoran saling independen. Efek kovariat ditaksir melalui vektor regresi β tanpa membuat suatu asumsi tentang fungsional dari *baseline hazard* dengan memaksimumkan fungsi kemungkinannya (*likelihood function*).

Metode kemungkinan maksimum digunakan untuk menaksir β yang dapat memaksimumkan peluang yang diperoleh dari waktu kegagalan yang diamati. Metoda kemungkinan maksimum marginal akan menghasilkan harga taksiran vektor regresi β yang sama. Oleh karena itu, penggunaan pendekatan kemungkinan maksimum akan digunakan untuk menaksir β dalam model regresi *Cox proportional hazard* dengan penambahan variabel terikat oleh waktu.

Misalkan $t_{(1)}, \dots, t_{(r)}$ adalah waktu peristiwa kematian dari data kegagalan berpasangan dan tidak dilakukan sensor, *r* adalah banyaknya waktu kegagalan. $t_{(j)}$ adalah waktu kematian pada saat ke-j. Diberikan pula x_j adalah kovariat ke-j yang berhubungan dengan individu yang memiliki waktu ketahanan $t_{(j)}$. $R\left(t_{(j)}\right)$ adalah himpunan dari semua individu yang masih berada di bawah pengamatan pada waktu sebelum $t_{(j)}$. Kemudian untuk setiap j, penaksir kemungkinan maksimum untuk kovariat dihasilkan melalui persamaan:

$L_j(\beta)$ = Probabilitas individu yang mati dengan kovariat x_j saat $t_{(j)}$ diberikan satu individu mati di $R_{(j)}$ saat $t_{(j)}$.

$$= \mathrm{P}\left(\text{individu mati dengan } x_j \text{ saat } t_{(j)} \middle| 1 \text{ individu mati di } R_{(j)} \text{ saat } t_{(j)}\right)$$

$$= \frac{\mathrm{P}\left(\text{individu mati dengan } x_j \text{ saat } t_{(j)} \middle| \text{individu di } R\left(t_{(j)}\right)\right)}{\mathrm{P}\left(\text{satu individu mati saat } t_{(j)} \middle| R\left(_{(j)}\right)\right)}$$

$$= \frac{h_0\left(t_{(j)}\right).\exp\left(x_j\beta\right)}{\sum_{l\in R\left(t_{(j)}\right)} h_0\left(t_{(j)}\right).\exp\left(x_l\beta\right)}$$

$$= \frac{\exp\left(x_j\beta\right)}{\sum_{l\in R\left(t_{(j)}\right)} \exp\left(x_l\beta\right)} \qquad \text{... (3.3.1)}$$

sehingga fungsi maksimum likelihood dapat dinotasikan sebagai berikut:

$$L(\beta) = \prod_{j=1}^{r} L_j(\beta)$$

$$= \prod_{j=1}^{r} \frac{\exp\left(x_j\beta\right)}{\sum_{l\in R\left(t_{(j)}\right)} \exp\left(x_l\beta\right)} \qquad \text{... (3.3.2)}$$

Pada umumnya vektor regresi β dapat ditaksir dengan memaksimalkan fungsi likelihood yang diperoleh dengan mempertimbangkan laju kegagalan yang terjadi antar waktu kegagalan. Dalam menaksir nilai-nilai 燿 yang memaksimumkan fungsi $L(\beta)$ dapat digunakan prinsip teorema dasar kalkulus yaitu dengan cara menentukan turunan pertama dan kedua dari $L(\beta)$ terhadap masing-masing β. Untuk lebih mempermudah penentuan turunan dari $L(\beta)$ dapat diambil bentuk logaritma dari $L(\beta)$ sehingga diperoleh persamaan sebagai berikut.

$$LogL(\beta) = \sum_{j=1}^{r} x_j\beta - \sum_{j=1}^{r} log\left[\sum_{l\in R\left(t_{(j)}\right)} \exp\left(x_l\beta\right)\right] \qquad \ldots (3.3.3)$$

Maka turunan pertama dari log $L(\beta)$ adalah

$$U_m(\beta) = \frac{\partial \log L(\beta)}{\partial \beta_m} = \sum_{j=1}^{r}\left\{x_{jm} - A_{j}(\beta)\right\} \qquad \ldots (3.3.4)$$

dimana $A_{jm}(\beta) = \dfrac{\sum_{l\in R\left(t_{(j)}\right)} x_{lm} \,.\, exp(x_l\beta)}{\sum_{l\in R\left(t_{(j)}\right)} exp(x_l\beta)}$

Turunan kedua dari log $L(\beta)$ adalah matriks I yang merupakan matriks informasi Fisher yang terdiri dari:

$$I_{mn}(\beta) = -\frac{\partial^2 L(\beta)}{\partial \beta_m \partial \beta_n} = \sum_{j=1}^{r} C_{(jmn)}(\beta) \qquad \ldots (3.3.5)$$

Dimana $C_{(jmn)}(\beta) = \left\{\dfrac{\sum_{l\in R\left(t_{(j)}\right)} x_{lm}x_{ln} \,.\, exp(x_l\beta)}{\sum_{l\in R\left(t_{(j)}\right)} exp(x_l\beta)}\right\} - A_{jm}(\beta)A_{jn}(\beta)$

Nilai-nilai β dapat diperoleh dengan memaksimalkan $L(\beta)$ yang dipecahkan dengan menggunakan metode Newton-Raphson (Lawless, 2003). Langkah-langkah yang dapat ditempuh agar dapat menggunakan metode Newton-Raphson adalah sebagai berikut:

1. Tentukan nilai awal taksiran parameter dalam model, $\beta^{(0)}$.
2. Tentukan vektor skor statistik ($U(\beta)$) dan matriks informasi ($I(\beta)$). Dalam hal ini, vektor statistik merupakan turunan pertama dari fungsi log-likelihood

terhadap parameter-parameternya atau $\frac{\partial \log L(\beta)}{\partial \beta} = U(\beta)$. Sedangkan matriks informasi merupakan turunan kedua dari fungsi log-likelihood terhadap parameter-parameternya atau $-\frac{\partial^2 L(\beta)}{\partial \beta^2} = I(\beta)$.

3. Untuk iterasi ke-m hitung:

   $$\underline{\beta}^{(m+1)} = \underline{\beta}^{(m)} - \left[I(\beta^{(m)})\right]^{-1} U(\beta^{(m)}), \qquad \text{untuk } m = 0, 1, 2, \ldots$$

4. Apabila selisih mutlak antara $\underline{\beta}^{(m)}$ dengan $\underline{\beta}^{(m+1)}$ mendekati nol maka proses iterasi berhenti. Tetapi jika $\left|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}\right| > 0{,}0001$, maka ulangi langkah ke-2 dan langkah ke-3 sampai memperoleh $\left|\underline{\beta}^{(m)} - \underline{\beta}^{(m+1)}\right| \leq 0{,}0001$.

Aplikasi dari metode diatas telah tersedia dalam sejumlah *software* diantaranya SPSS, SPIDA, SAS, dan S-PLUS. Dalam tugas akhir ini penulis menggunakan software SPSS versi 16.0 dalam konteks *survival analysis* melalui model *cox with time dependent covariates*.

3.4 Pengujian Hipotesis

3.4.1 Pengujian Variabel Bebas oleh Waktu

3.4.1.1 Pengujian Signifikansi Parameter

Uji signifikansi parameter ini dikenal dengan uji rasio likelihood. Pengujian ini digunakan untuk mengetahui apakah model regresi *Cox proportional hazard* cocok dengan data secara signifikan atau tidak (keberartian model). Perumusan hipotesis untuk uji rasio likelihood adalah sebagai berikut:

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

H_1: paling sedikit satu tanda sama dengan tidak berlaku (paling sedikit satu parameter berbeda secara signifikan dengan nol)

Kriteria uji:

Tolak H_0 jika $X_{LR}^2 = (-2logL(\beta_0) - \left(-2logL(\widehat{\beta})\right) \geq \chi_{tabel}^2 = \chi_{(d)}^2$

Dengan derajat kebebasan adalah selisih banyaknya parameter antara model sekarang dengan model sebelumnya dan taraf kepercayaan yang diinginkan yaitu sebesar 95% sehingga taksiran untuk $\widehat{\beta}$ dapat dihitung.

3.4.1.2 Pengujian Keberartian Koefisien

Untuk uji keberartian koefisien regresi digunakan uji Wald dimana dengan pengujian ini variabel bebas dapat dikeluarkan dari model ketika efeknya tidak signifikan. Perumusan hipotesis untuk uji Wald adalah sebagai berikut:

$H_0 : \beta = \beta_i = 0$

$H_1 : \beta \neq \beta_i \neq 0$ dengan i = 1, 2, …, p

Kriteria uji:

Tolak H_0 jika $X_W^2 = \left[\frac{b}{SE(b)}\right]^2 \geq \chi_{tabel}^2 = \chi_{(d)}^2$

Dengan derajat kebebasan adalah 1.

3.4.2 Pengujian Variabel Terikat oleh Waktu

Sama halnya seperti pengujian variabel bebas oleh waktu, pengujian signifikansi parameter dan keberartian koefisien pada variabel terikat oleh waktu juga menggunakan uji rasio likelihood. Berikut adalah uraian metode pengujiannya.

3.4.2.1 Pengujian Signifikansi Parameter

Perumusan hipotesis untuk uji signifikansi parameter adalah sebagai berikut:

H_0: $\delta_1 = \delta_2 = ... = \delta_p = 0$

H_1: paling sedikit satu tanda sama dengan tidak berlaku (paling sedikit satu parameter berbeda secara signifikan dengan nol)

Kriteria uji:

Tolak H_0 jika $X_{LR}^2 = (-2logL(\beta_0) - \left(-2logL(\widehat{\beta})\right) \geq \chi_{tabel}^2 = \chi_{(d)}^2$

Taraf kepercayaan yang diinginkan yaitu sebesar 95% sehingga taksiran untuk $\widehat{\beta}$ dapat diperoleh.

3.4.2.2 Pengujian Keberartian Koefisien

Perumusan hipotesis untuk uji Wald adalah sebagai berikut:

H_0 : $\delta = \delta_i = 0$

H_1 : $\delta \neq \delta_i \neq 0$ dengan i = 1, 2, …, p

Kriteria uji:

Tolak H_0 jika $X_W^2 = \left[\frac{b}{SE(b)}\right]^2 \geq \chi_{tabel}^2 = \chi_{(d)}^2$

Dengan derajat kebebasan adalah 1.

3.5 Penentuan Nilai Baseline Hazard

Baseline hazard dapat dikatakan sebagai nilai *hazard* jika kovariat-kovariat yang ada dianggap tidak memberikan pengaruhnya terhadap bentuk ketahanan. Nilai *baseline hazard* memperlihatkan bentuk laju waktu ketahanan masa hidup suatu individu apabila semua pengaruh-pengaruh dari semua kovariat dianggap nol. Terdapat dua cara perhitungan yang dapat digunakan untuk menaksir model nilai *baseline hazard* ($h_0(t)$) (Permana, 1999) yaitu sebagai berikut.

1. Penggunaan asumsi dari distribusi parametrik.
2. Tidak ditentukannya asumsi khusus untuk bentuk distribusi.

   Cara kedua ini lebih banyak digunakan karena untuk menentukan distribusi khusus dari laju kegagalan yang dipengaruhi oleh kovariat-kovariat sering kali membingungkan. Alasan lainnya dikarenakan pendekatan kedua ini lebih sering digunakan karena bentuknya lebih umum.

Apabila taksiran dari vektor regresi $\widehat{\beta}$ telah diperoleh maka nilai *baseline hazard* dan fungsi *baseline hazard* kumulatif dapat dihitung (Lawless, 2003) melalui persamaan sebagai berikut.

1. *Baseline hazard*.

$$h_0(t) = \frac{d_i}{\sum_{l \in R\left(t_{(j)}\right)} \exp\left(x_l \beta\right)} \qquad \text{…(3.4.1)}$$

2. Fungsi *baseline hazard* kumulatif.

$$H_0(t) = \sum_{i:t_{(j)}<t} \frac{d_i}{\sum_{l \in R\left(t_{(j)}\right)} \exp\left(x_l \beta\right)} \qquad \text{…(3.4.2)}$$