BAB II

TINJAUAN PUSTAKA

2.1 Konsep Dasar Distribusi *Survival*

Data *survival* adalah data lamanya individu-individu atau unit-unit dari suatu populasi menjalankan fungsinya dengan baik sampai kematian individu-individu tersebut. Dalam mempelajari penerapan data survival, terlebih dahulu harus diketahui konsep-konsep statistik pada distribusi *survival* (Sari, 2011).

Analisis data tahan hidup (*survival analysis*) adalah suatu metode untuk menganalisis yang berhubungan dengan waktu, mulai dari *time origin* atau *start-point* sampai dengan terjadinya suatu kejadian khusus atau *end point.* Di bidang kesehatan data diperoleh dari pengamatan terhadap pasien yang diamati dan dicatat waktu terjadinya *event* dari setiap individu. *Event* yang dimaksut dapat berupa kematian, kambuhnya penyakit baru, atau kesembuhan. (Collet, 2013)

Menurut Kleinbaum dan Klein (2005), dalam menentukan waktu survival T, terdapat tiga elemen yang perlu diperhatikan, yaitu:

1. *Time Origin or Starting Point* (titik awal) adalah waktu dimulainya suatu penelitian. Titik awal pada penelitian ini adalah tangga masuk pasien rawat inap DBD di Rumah Sakit.
2. *Ending Event of interest* (kejadian akhir) adalah kejadian yang menjadi inti dari penelitian. Titik akhir yang dimaksud pada penelitian adalah tanggal dimana pasien rawat inap DBD yang dinyatakan keluar dari Rumah Sakit

dalam keadaan sembuh.

3. *Measurement scale for the passage of time* (skala ukuran untuk berlalunya waktu). Dalam penelitian ini skala yang digunakan adalah lama pasien DBD yang rawat inap di Rumah Sakit dalam satuan hari.

2.2 Analisis *Survival*

Analisis *survival* merupakan suatu metode statistik yang berkaitan dengan waktu, yaitu dimulai dari *time origin* atau *start point* sampai pada suatu kejadian khusus (*failure event/end point*). Salah satu analisis survival yang digunakan adalah regresi *cox*, yaitu suatu regresi yang digunakan untuk analisis data dengan variabel dependennya berupa waktu *survival*. Selisih waktu mulai dilakukannya pengamatan sampai waktu terjadinya kematian atau disebut data tidak tersensor (*uncensored data*). Jika waktu kematiannya tidak diketahui, maka dipakai selisih waktu dari mulai dilakukannya pengamatan sampai waktu akhir penelitian disebut data tersensor (*censored data*).

Beberapa kemungkinan penyebab kematian data tersensor adalah masa penelitian berakhir sementara observasi masih belum mencapai failure. Kematian karena sebab tertentu disebut failure event. Suatu pengamatan waktu s*urvival* tidak sampai pada *failure event*, penyebabnya adalah *Lost of follow up, drop out*, dan *termination of the study*.

Fungsi *survival* digunakan untuk menyatakan probabilitas suatu individu bertahan dari waktu mula-mula sampai waktu t. Waktu *survival* dilambangkan dengan T yang merupakan variabel random dan mempunyai fungsi distribusi peluang f (t). Fungsi *survival* S(t), didefinisikan sebagai probabilitas bahwa waktu survival lebih besar atau sama dengan t sehingga:

$$S(t : \mu,\sigma) = 1 - \phi\left(\frac{\ln(t)\ -\ \mu}{\sigma}\right)$$

(2.1)

Fungsi *hazard* menyatakan proporsi atau laju kematian seketika suatu individu yang *survive* sampai waktu ke-*t*. Berikut adalah fungsi hazard tanpa pengaruh variabel bebas yang biasa disebut dengan fungsi *baseline hazard* :

$$h(t : \mu,\ \sigma) = \frac{\phi\left(\frac{\ln(t)\ -\ \mu}{\sigma}\right)}{\left[1\ -\ \phi\ \left(\frac{\ln(t)\ -\ \mu}{\sigma}\right)\right]\sigma(t)}$$

(2.2)

2.3 Metode *Kaplan- Meier*

Metode *Kaplan-Meier* merupakan salah satu metode nonparametrik yang dapat digunakan untuk menduga fungsi daya tahan tanpa mengikutsertakan peubah penjelas. Metode ini tidak memerlukan asumsi sebaran tertentu. Metode Kaplan-Meier mengelompokkan data ke dalam suatu selang, dalam setiap selang memuat satu kejadian. Misal *t1, t2, ..., tn* adalah durasi daya tahan dari *n* individu dalam pengamatan. Durasi daya tahan tersebut diurutkan dari yang terkecil hingga yang terbesar. Misal t_k adalah durasi daya tahan ke-k, dengan $t_1 < t_2 < \dots < t_D$ *D* adalah banyaknya durasi daya tahan yang berbeda dengan $D \leq n$. Peluang kejadian dalam setiap selang *k* diduga dengan $\frac{d_k}{n_k}$ dan peluang bertahannya diduga dengan

$\hat{P}_k = \frac{n_k - d_k}{n_k}$ dengan d_k adalah banyaknya kejadian pada setiap selang *k*, dan nk adalah banyaknya individupada awal selang *k*. Penduga fungsi daya tahan dengan menggunakan metode *Kaplan-Meier* adalah sebagai berikut :

$$\hat{S}(k) = \hat{P}_1 \times \hat{P}_2 \times \hat{P}_3 \times \ldots \times \hat{P}_k$$

$$= \prod_{k=1}^{D} \hat{P}_k \quad (2.3)$$

2.4 Uji Perbedaan antar Kelompok Data *Survival*

Uji *log rank* digunakan untuk menguji apakah secara statistik terdapat perbedaan pada kurva *survival* Kaplan-Meier antara dua kelompok data atau lebih. Uji *log rank* membandingkan jumlah kejadian hasil observasi pada masing-masing kelompok data dengan nilai ekspektasinya (Kleinbaum dan Klein, 2012). Hipotesis yang digunakan pada uji *log rank* untuk dua atau lebih kelompok adalah sebagai berikut:

H_0: Tidak ada perbedaan pada kurva *survival* antara grup yang berbeda

H_1: Minimal terdapat satu perbedaan pada kurva *survival* antara grup yang berbeda

Statistik uji pada uji *log rank* adalah

$$X^2 \approx \sum_{i=1}^{G} \frac{(O_i - E_i)^2}{E_i}$$

(2.4)

Dimana

$$O_i - E_i = \sum_{j=1}^{n} \sum_{i=1}^{G} (m_{ij} - e_{ij}) \text{dan } e_{ij} = \left(\frac{n_{ij}}{\sum_{j=1}^{n} \sum_{i=1}^{G} n_{ij}}\right)\left(\sum_{j=1}^{n} \sum_{i=1}^{G} m_{ij}\right)$$

(2.5)

Keterangan

O_i : nilai observasi individu grup ke-*i*

E_i : nilai ekspektasi individu grup ke-*i*

m_{ij} : jumlah subjek yang gagal dalam grup ke-*i* pada waktu □(□)

n_{ij} : jumlah subjek yang beresiko gagal seketika pada grup ke-*i* sebelum waktu □(□)

e_{ij} : nilai ekspektasi dalam grup ke-*i* pada waktu □(□)

G : banyak grup

2.5 Regresi *Cox Proportional Hazard (Cox PH)*

Pada analisis *survival* terdapat dua model, yaitu parametrik dan semiparametrik. Model parametrik antara lain model *Weibull* yang berdistribusi *weilbull* dan model *eksponensial* yang berdistribusi *eksponensial* (Kleinbaum dan Klein, 2005).

Menurut Lee dan Wang (2003), model regresi *Cox Proportional Hazard (Cox PH)* merupakan model berdistribusi semiparametrik karena model *Cox PH* tidak memerlukan informasi tentang distribusi yang mendasari waktu survival dan parameter regresi dapat diestimasi dari model. Kenyataannya, data yang diperoleh tidak dapat memberikan informasi distribusi waktu *survival*, sehingga bentuk $h_0(t)$ dari fungsi *hazard* dasar juga tidak dapat diketahui. Model semiparametrik lebih sering digunakan karena bentuk fungsional $h_0(t)$ tidak diketahui, tapi model *Cox PH* ini tetap dapat memberikan informasi berupa *hazard ratio (HR)* yang tidak bergantung pada $h_0(t)$*HR* didefinisikan sebagai rasio dari *hazard rate* satu individu dengan *hazard rate* dari individu lain.

Asumsi *PH* yaitu jika sebuah garis pada kurva *survival* (antar kelompok) tidak saling berpotongan. Asumsi *PH* sangat penting dalam analisis *survival*. Analisis yang dilakukan pada suatu fungsi yang memenuhi asumsi *PH* berbeda dengan analisis yang dilakukan pada fungsi *survival* yang tidak memenuhi asumsi *Cox PH* (Kleinbaum DG dan Klein M., 2005).

Ada tiga jenis pengecekan asumsi *Proportional Hazard* menurut Dahlan, 2013 yaitu:

1. Garis *survival* pada kurva *Kapplan – Meier* tidak saling berpotongan
2. Garis *survival* pada *ln-ln survival* tidak saling berpotongan
3. Uji *globaltest*

Dari ketiga jenis pengecekan asumsi Proportional Hazard, pada penelitian ini menggunakan pengecekan asumsi dengan uji global test. Jika asumsi telah terpenuhi, maka model Cox PH dapa dibentuk. Salah satu tujuan model Cox PH adalah untuk memodelkan hubungan antara waktu survival dengan peubah-peubah yang diduga mempengaruhi waktu survival.

Pengecekan asumsi PH pada data dilakukan sebagai berikut:

Hipotesis dari uji global test:

H_o = Data memenuhi asumsi PH

H_1 = Data tidak memenuhi asumsi *PH*

Signifikansi α : 0,05

H_o ditolak jika *p- value* <α

Model *Cox PH* dapat dituliskan sebagai berikut.

$$h_i(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)$$

(2.6)

$h_i(t)$ = fungsi kegagalan individu ke-i

$h_0(t)$ = fungsi kegagalan dasar (fungsi *Hazard*)

β_i = nilai peubah ke-i , dengan i = 1,2, . . . , n

x_i = koefisien regresi ke-i, dengan i = 1,2,. . ., n

(Kleinbaum dan Klein, 2005)

2.6 Pendugaan Parameter

Pendugaan parameter *β* dapat dicari melauli fungsi kemungkinan parsial yang didasarkan pada peluang bersyarat. Misalkan t_k adalah waktu kejadian ke- k, dengan $t_1 < t_2 < \dots t_D$, *D* adalah banyaknya waktu kejadian yang berbeda. $X_{(k)j}$ adalah peubah penjelas ke- j dari individu dengan waktu kejadian t_k dan $R(t_k)$ adalah himpunan individu yang memiliki resiko pada waktu t_k. Peluang *cox* individu ke- i peubah *X* mengalami kejadian pada waktu t_k adalah:

$$l(\beta) = \frac{P(X_k \text{ mengalami kejadian } t_k)}{P(R(t_k \text{ mengalami kejadian } t_k))} = \frac{h_i(t_k | X_{(k)j})}{\sum_{m \in R(t_k)} h_i(t_k | X_{(m)j})}$$

$$\frac{\exp[\sum_{j=1}^{p} \beta_j X_{(k)j}]}{\sum_{m \in R(t_k)} \exp[\sum_{j=1}^{p} \beta_j X_{(m)j}]} \quad (2.7)$$

2.7 Asumsi *Hazard* Proportional

Regresi *hazard* proporsional memiliki asumsi yang cukup kuat yaitu memiliki *hazard* yang bersifat proporsional antara satu individu dan individu lainnya. Dengan kata lain, rasio tingkat *hazard* antar dua kelompok individu (misal individu *i* dan *i'*) harus konstan sepanjang waktu. (Therneau dan Grambsch 2000).

$$\frac{h(t|x_i)}{h(t|x_{i'})} = \frac{h_0(t).\exp(\beta' x_i)}{h_0(t).\exp(\beta' x_{i'})} = \frac{\exp(\beta' x_i)}{\exp(\beta' x_{i'})}$$

(2.8)

Kleinbaum dan Klein (2012) menyebutkan bahwa ada beberapa pendekatan yang dapat dilakukan untuk pemeriksaan asumsi *hazard* proporsional, yaitu dengan pendekatan grafik –ln(-ln S(t)) dan pengujian kesesuaian model (goodness of-fit test) yaitu residual Schoenfeld. Residual Schoenfeld didasarkan pada kontribusi individu dengan turunan dari log kemungkinan parsial (Hosmer dan Lemeshow 1999).

$$\frac{\partial L_p(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} c_i \left\{ x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}} \right\}$$

(2.9)

Dimana $\bar{x}_{w_i k} = \frac{\sum_{j \in R(t_i)} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}}$

Hosmer dan Lemeshow (1999) menunjukkan, penduga sisaan Schoenfeld untuk subjek ke-i pada kovariat ke-k kemudian diperoleh dengan mengganti penduga parsial likelihood koefisien $\hat{\beta}$,

$$Y_{S_{ik}} = c_i(x_{ik} - \bar{x}_{w_{ik}})$$

(2.10)

Kemudian, dibuat peubah peringkat urutan waktu kegagalan. Subjek yang terjadi pertama (awal) kejadian mendapat nilai 1, berikutnya mendapat nilai 2, dan seterusnya. Setelah itu, uji korelasi antara sisaan Schoenfeld dan peringkat urutan kegagalan. H0 yang digunakan bahwa korelasi antara sisaan Schoenfeld dan peringkat waktu kegagalan adalah nol (ρ = 0). Statistik uji yang digunakan adalah uji-t yaitu

$$t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$$

(2.11)

Besar α yang digunakan adalah 0.05. Jika *p-value* kurang dari α = 0.05, maka H0 ditolak, artinya terdapat korelasi yang menunjukkan ada hubungan peubah penjelas dengan waktu daya tahan sehingga tidak memenuhi asumsi (Kleinbaum dan Klein 2012).

2.8 Pengujian Penduga Parameter

Pengujian penduga parameter dilakukan untuk memeriksa peranan peubah-peubah penjelas dalam model Cox. Pengujian parameter model (⍰) secara simultan dilakukan menggunakan statistik uji-*G*, sedangkan pengujian secara parsial menggunakan statistik uji Wald. Statistik uji-*G* adalah uji rasio kemungkinan yang digunakan untuk menguji peranan variabel penjelas di dalam model secara bersama-sama. Hipotesis yang diuji adalah:

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

$$H_1 : \text{Minimal ada satu } \beta_j \neq 0,\ j = 1, 2, \dots, p.$$

Rumus umum statistik uji-G sebagai berikut:

$$G = -2 \ln\left[\frac{L_0}{L_p}\right] \tag{2.12}$$

dengan L_0 adalah fungsi kemungkinan tanpa peubah penjelas dan L_p adalah fungsi kemungkinan dengan *p* peubah penjelas. Statistik uji *G* menyebar khi-kuadrat dengan derajat bebas *p*. Apabila nilai *G* lebih besar dari nilai tabel khi-kuadrat pada α=5% mengindikasikan bahwa terdapat minimal satu peubah penjelas yang pengaruhnya signifikan terhadap durasi daya tahan debitur.

Statistik uji Wald digunakan untuk menguji penduga parameter β_j secara parsial. Hipotesis yang diuji adalah:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0,\ j = 1, 2, \dots, p.$$

Rumus umum statistik uji Wald sebagai berikut:

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{2.13}$$

Secara teori, statistik uji Wald ini mengikuti sebaran khi-kuadrat. Jika $W > \chi^2_{\alpha,1}$ maka H_0 ditolak pada taraf nyata α = 5% (Collet 2003).

2.9 Regresi *Cox Proportional Hazard*

Regresi Cox didasarkan pada beberapa asumsi. Salah satunya adalah asumsi *proportional hazard* dimana merupakan hasil dari rumus sebagai berikut:

$$h_i(t) = h_0(t)e^{\beta_1\beta_1 + \beta_2\beta_2 + \ldots + \beta_i\beta_i + \beta_k\beta_k} \quad (2.14)$$

Dimana:

t = waktu hingga suatu kejadian tertentu

$h_0(t)$ = baseline hazard

β_i = koefisien regresi

X_i = variabel bebas, $i = 1, 2, \ldots, k$

Asumsi *proportional hazard* menyatakan bahwa *hazard ratio* untuk dua subjek dari kelompok yang berbeda hanya tergantung pada nilai kovariat dan tidak tergantung dengan waktu. Dengan kata lain, *hazard ratio* konstan sepanjang waktu yang berarti bahwa efek dari kovariat pada *hazard level* adalah sama untuk semua waktu. Ada beberapa pendapat berkaitan dengan pentingnya asumsi ini. Beberapa peneliti berpendapat tidak masalah dengan pelanggaran asumsi ini, yang dapat diartikan sebagai efek rata-rata dari semua waktu yang diobservasi (Allison, 2010). Sedangkan peneliti lain menyatakan pentingnya asumsi ini (Hosmer dan Lemeshow, 2008) dan menyarankan beberapa model jika *hazard ratio* tidak konstan sepanjang waktu untuk beberapa kovariat.

2.10 Model Regresi *Stratified Cox*

Model s*tratified* Cox merupakan perluasan dari model Cox *proportional hazard* untuk mengatasi variabel bebas yang tidak memenuhi asumsi proporsional hazard. Asumsi *proportional 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). Modifikasi dilakukan dengan menstratifikasi variable bebas yang tidak memenuhi asumsi *proportional hazard*. Variabel bebas yang memenuhi asumsi *proportional hazard* masuk ke dalam model, sedangkan variabel bebas yang tidak memenuhi asumsi, yang sedang distratifikasi, tidak masuk dalam model (Kleinbaum dan Klein, 2012).

Dalam model s*tratified* Cox diasumsikan terdapat sebanyak *p* variable bebas. Sebanyak *k* variabel bebas diantaranya memenuhi asumsi *proportional hazard* dinotasikan □1, □2, … , □□ dengan □ < □. Variabel bebas yang tidak memenuhi asumsi *proportional hazard* sebanyak *m* yang diperoleh dari □ − □ = □ yaitu □□+1, □□+2, … , □□ yang dinotasikan □1, □2, … , □□ .

$$X_{k+1} \to Z_1 : X_{k+2} \to Z_2 : \dots : X_p \to Z_m$$

Variabel yang tidak memenuhi asumsi *proportional hazard* □□ dengan *i* = 1, ..., *m* dikeluarkan dari model *cox* untuk dilakukan stratifikasi terhadap variabel tersebut sehingga diperoleh variabel stratifikasi □□. Variabel bebas yang memenuhi asumsi *proportional hazard* akan masuk ke dalam model s*tratified cox*. Meskipun begitu variabel bebas yang dikeluarkan dari model tetap memiliki peran dan dengan dilakukan stratifikasi variabel akan terlihat kontribusi masing-masing variable bebas tersebut dalam strata yang berbeda. Langkah pertama untuk membentuk model regresi s*tratified* Cox adalah menguji interaksi pada model. Untuk menguji ada tidaknya interaksi pada model s*tratified* Cox digunakan uji *likelihood ratio* (LR) yaitu dengan membandingkan statistik log *likelihood* untuk model interaksi dan model tanpa interaksi (Kleinbaum & Klein, 2012). Hipotesis dari uji *likelihood ratio* (LR) adalah sebagai berikut.

H_0: Tidak ada interaksi antara variabel stratifikasi dengan variabel independen yang masuk dalam model

H_1: Terdapat interaksi antara variabel stratifikasi dengan variabel independen yang masuk dalam model

Statistik uji:

$$LR = -2\ln L_R - (-2\ln L_F) \sim X^2_{p(k^*-1)}$$

(2.15)

Dimana

R = Model tanpa interaksi

F = Model dengan interaksi

Tolak H_0 jika $LR > X^2_{p(k^*-1)}$ atau < α.

Menurut Kleinbaum dan Klein (2012) bentuk umum fungsi *hazard* dari model *stratified cox* tanpa interaksi adalah sebagi berikut:

$$h_s(t,X) = h_0(t)\exp[\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k]$$

(2.16)

Dengan

S = strata yang didefinisikan dari Z. S= 1,2,...m

$h_{0s}(t)$ = fungsi dasar hazard untuk setiap strata

$\beta_1, \beta_2, \dots \beta_k,$ = parameter regresi