

## MODEL REGRESI COX PROPORSIONAL HAZARD PADA DATA KETAHANAN HIDUP

Tuan Hanni<sup>1</sup>, Triastuti Wuryandari<sup>2</sup>

<sup>1</sup>Alumni Jurusan Statistika FSM UNDIP

<sup>2</sup>Staf Pengajar Jurusan Statistika FSM UNDIP

### Abstract

A lot of events occurred in daily life are connected with survival time, for example a time interval that measure the failure of a product, time duration which is needed to recover from disease, the back pain recurred after treatment. Data about survival time duration of an event is called survival data. Survival data can not be observed completely that is called as censored data. Cox proportional hazard model is employed to analyze and determine the survival rate from censored data affected one or more explanatory variables. This model assumed that the hazard rate of group is proportional to the hazard rate of another group. In the paper, wants to the factor that affect the survival of patient with cervical cancer. From the result of data processing that affect are age and stadium with cox proportionl hazard model is  $h_i(t) = \exp(-1.848U1_i - 1.584U2_i - 3.255S2_i - 2.108S3_i) h_0(t)$

**Keywords :** Cox Proportional Hazard, Survival Rate, Hazard Rate, Cervical Cancer

### 1. Pendahuluan

Analisis survival/analisis data ketahanan hidup adalah suatu metode yang berhubungan dengan waktu, mulai dari awal pengamatan sampai terjadinya kejadian khusus. Analisis survival memerlukan data yang merupakan waktu ketahanan dari suatu individu. Di bidang kesehatan data ini diperoleh dari suatu pengamatan terhadap sekelompok atau beberapa kelompok individu yang diamati waktu terjadinya kegagalan dari setiap individu<sup>[4]</sup>. Kegagalan yang dimaksud antara lain adalah kematian karena penyakit tertentu, kambuhnya suatu penyakit atau munculnya penyakit baru.

Ada beberapa teori yang pernah membahas tentang *survival analysis* diantaranya adalah *Kaplan-meier* dan *Cox*. Model regresi *cox proportional hazard* dapat menjelaskan pengaruh faktor independen dalam suatu kejadian. Pada penelitian ini, ingin diketahui faktor-faktor yang mempengaruhi ketahanan hidup penderita kanker leher rahim. Populasi dalam penelitian ini adalah semua penderita kanker leher rahim yang pernah menjalani perawatan di rumah sakit Dr. Cipto Mangunkusumo Jakarta periode Januari 1997 sampai dengan September 2002, sedangkan sampel yang diambil adalah penderita kanker leher rahim yang bertempat tinggal di Jakarta. 80 orang. Data diperoleh dari data rekam medik rumah sakit. Sumber: *Ketahanan Hidup Penderita Kanker Serviks di Rumah Sakit Dr. Cipto Mangunkusumo Jakarta*. Indones J. Obstet Gynecol 21(3): 182-190<sup>[12]</sup>.

Permasalahan dibatasi pada pembentukan model regresi *cox proportional hazard* dengan faktor-faktor yang mempengaruhi ketahanan hidup pasien penderita kanker leher rahim. Data yang digunakan untuk aplikasi kasus bersumber dari data ketahanan hidup penderita kanker leher rahim dengan penyensoran tipe III. Penaksiri parameternya menggunakan metode *Maximum Likelihood Estimation* (MLE). Pengolahan datanya menggunakan *software* statistik SPSS 16.

## 2. Konsep dasar

Data ketahanan hidup merupakan interval waktu yang diamati dari suatu individu saat pertama kali masuk pengamatan sampai dengan individu tersebut gagal atau mati. Misalnya interval waktu yang mengukur kerusakan suatu produk, matinya suatu makhluk hidup karena mengidap suatu penyakit dan interval waktu kambuhnya suatu penyakit. Fungsi-fungsi pada distribusi waktu hidup merupakan suatu fungsi yang menggunakan variabel random waktu hidup. Variabel random waktu tahan hidup dinotasikan dengan huruf  $T$ . Selanjutnya akan dibentuk suatu distribusi yang disebut distribusi waktu hidup. Distribusi waktu hidup terdiri dari tiga fungsi, yaitu:

1. Fungsi tahan hidup (fungsi *survival*)
2. Fungsi densitas peluang
3. Fungsi kegagalan (fungsi *hazard*)

Ketiga fungsi tersebut ekuivalen secara matematik, artinya jika salah satu dari ketiga fungsi tersebut diketahui maka fungsi yang lain dapat ditentukan<sup>[11]</sup>.

### 2.1. Fungsi Tahan Hidup (Fungsi *Survival*)

Fungsi ketahanan hidup adalah peluang suatu individu akan beroperasi dengan baik untuk periode waktu tertentu di bawah kondisi operasi yang ditentukan. Ketahanan hidup dapat digunakan sebagai pengukur keberhasilan suatu sistem dalam menjalankan fungsinya dengan baik. Misalkan  $n_0$  adalah banyaknya individu penderita penyakit tertentu. Selama interval waktu  $(t, t + \Delta t)$ , diamati banyaknya individu yang gagal akibat penyakit tersebut  $n_f(t)$ . Banyaknya individu bertahan hidup  $n_s(t)$ , sehingga  $[n_f(t) + n_s(t) = n_0]$ . Jika ketahanan hidup didefinisikan sebagai fungsi peluang kumulatif seorang pasien bertahan hidup lebih dari waktu  $t$ , dengan  $t > 0$ , maka fungsi tahan hidup  $S(t)$  adalah

$$S(t) = \frac{n_s(t)}{n_s(t) + n_f(t)} = \frac{n_s(t)}{n_0} \quad (1)$$

$$\begin{aligned} S(t) &= P(\text{individu hidup lebih dari waktu } t) \\ &= P(T > t) = 1 - P(\text{individu gagal sebelum waktu } t) = 1 - P(T \leq t) \end{aligned} \quad (2)$$

Jika  $f(t)$  merupakan fungsi densitas peluang dari  $T$  maka fungsi distribusi kumulatif  $F(t)$

$$F(t) = P(T \leq t) = \int_0^t f(x) dx \quad (3)$$

$F(t)$  merupakan komplemen dari  $S(t)$ , sehingga

$$S(t) + F(t) = 1 \quad \text{atau} \quad S(t) = 1 - F(t) \quad (4)$$

Berdasarkan persamaan (3) dan (4) maka fungsi tahan hidup  $S(t)$  dapat ditentukan

$$S(t) = 1 - F(t) = 1 - \int_0^t f(x) dx \quad (5)$$

Dengan menurunkan Persamaan (5) terhadap  $t$  diperoleh

$$\frac{dS(t)}{dt} = -f(t) \quad (6)$$

Menurut Miller, 1998, fungsi tahan hidup  $S(t)$  adalah fungsi yang tidak naik atau monoton turun dengan sifat:

1.  $S(t) = 1$  untuk  $t = 0$ , artinya peluang individu untuk hidup pada saat  $t = 0$  adalah 1
2.  $S(t) = 0$  untuk  $t = \infty$ , artinya peluang individu untuk hidup pada saat  $t = \infty$  adalah 0

## 2.2. Fungsi Densitas Peluang

Waktu tahan hidup  $T$  mempunyai fungsi densitas peluang yang dinotasikan dengan  $f(t)$  dan didefinisikan sebagai peluang kegagalan suatu individu pada interval  $(t, t + \Delta t)$  per satuan waktu. Fungsi densitas peluang dinyatakan sebagai

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[ \frac{P(\text{objek gagal pada } (t, t + \Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[ \frac{P(t < T < t + \Delta t)}{\Delta t} \right] \quad (7)$$

## 2.3. Fungsi Kegagalan (Fungsi Hazard)

Fungsi kegagalan dari waktu tahan hidup  $T$  dinotasikan dengan  $h(t)$  dan didefinisikan sebagai peluang suatu individu gagal di dalam interval waktu  $(t, t + \Delta t)$  dengan diketahui bahwa individu tersebut telah hidup selama waktu  $t$ . Fungsi keagalannya dinyatakan dengan

$$h(t) = \lim_{\Delta t \rightarrow 0} \left[ \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t} \right] \quad (8)$$

Fungsi kegagalan juga dapat dinyatakan dengan menggunakan fungsi distribusi kumulatif  $F(t)$  dan fungsi densitas peluang  $f(t)$ , yaitu

$$h(t) = \frac{f(t)}{1 - F(t)}, \text{ dengan syarat } F(t) \neq 1 \quad (9)$$

Selanjutnya untuk fungsi kegagalan kumulatif didefinisikan dengan peluang kegagalan dari interval 0 hingga  $t$  yang dinyatakan dengan

$$H(t) = \int_0^t h(x) dx \quad (10)$$

## 2.4. Hubungan antara Fungsi-Fungsi Tahan Hidup

Ketiga fungsi tahan hidup yaitu fungsi tahan hidup  $S(t)$ , fungsi densitas peluang  $f(t)$  dan fungsi kegagalan  $h(t)$ , secara matematik mempunyai hubungan satu sama lain. Oleh karena itu, jika salah satu fungsi diketahui maka fungsi yang lain juga dapat ditentukan.

1. Dari persamaan (4) dan persamaan (9) diperoleh

$$h(t) = \frac{f(t)}{S(t)} \quad (11)$$

2. Jika fungsi densitas peluang adalah turunan pertama fungsi distribusi kumulatif, maka

$$f(t) = \frac{d}{dt} [1 - S(t)] = -\frac{dS(t)}{dt} = -S'(t) \quad (12)$$

3. Persamaan (12) dapat disubstitusikan ke Persamaan (11), dan diperoleh

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{\frac{dS(t)}{dt}}{S(t)} = -\frac{1}{S(t)} \frac{d}{dt} (S(t)) = -\frac{d}{dt} \log S(t) \quad (13)$$

4. Integral dari Persamaan (13) dari nol sampai  $t$  dan dengan menggunakan

$$S(0) = 1, \text{ diperoleh persamaan: } \int_0^t h(x) dx = -\log S(t) \text{ atau } H(t) = -\log S(t)$$

$$\text{maka } S(t) \text{ dapat juga ditentukan dengan } S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(x) dx\right] \quad (14)$$

5. Dari Persamaan (11) dan (14) akan diperoleh

$$f(t) = h(t) \exp[-H(t)] \quad (15)$$

## 2.5. Tipe-Tipe Penyensoran

Di dalam data tahan hidup, dikenal istilah penyensoran. Tujuan penyensoran adalah memperpendek waktu percobaan karena dalam mengukur waktu kegagalan atau kematian individu, kadang-kadang diperlukan waktu yang lama dan biaya yang tidak sedikit. Oleh karena itu, penyensoran biasanya dilakukan pada data tahan hidup. Observasi tidak tersensor adalah waktu tahan hidup yang dicatat dari individu yang mati selama waktu percobaan, yaitu waktu dari awal percobaan hingga mengalami kematian. Untuk observasi tersensor merupakan waktu tahan hidup suatu individu yang tidak diketahui secara pasti, dengan kata lain dicatat sedikitnya selama waktu percobaan<sup>[9]</sup>.

Dalam uji tahan hidup dibedakan tiga tipe penyensoran, yaitu:

### - Sensor Tipe I

Sensor tipe I adalah tipe penyensoran dimana percobaan akan dihentikan setelah mencapai waktu  $T$  yang telah ditentukan untuk mengakhiri semua  $n$  individu yang masuk pada waktu yang sama. Berakhirnya waktu uji  $T$  menjelaskan waktu sensor uji. Dengan kata lain, jika tidak terdapat individu yang hilang secara tiba-tiba maka waktu tahan hidup observasi tersensor sama dengan lama waktu pengamatan<sup>[9]</sup>.

### - Sensor Tipe II

Sensor tipe II adalah tipe penyensoran dimana sampel ke- $r$  merupakan observasi terkecil dalam sampel random berukuran  $n$  ( $1 \leq r \leq n$ ). Dari total sampel berukuran  $n$ , dengan  $n$  berlanjut sampai mati atau gagal maka percobaan akan dihentikan sampai  $r$  dari  $n$  unit mengalami kematian. Semua unit uji  $n$  masuk pada waktu yang sama<sup>[9]</sup>.

### - Sensor Tipe III

Pada sensor tipe III, individu atau unit uji masuk ke dalam percobaan pada waktu yang berlainan selama periode waktu tertentu. Beberapa unit uji mungkin gagal atau mati sebelum pengamatan berakhir sehingga waktu tahan hidupnya dapat diketahui secara pasti. Kemungkinan kedua adalah unit uji keluar sebelum pengamatan berakhir, atau kemungkinan ketiga adalah unit uji tetap hidup sampai batas waktu terakhir pengamatan<sup>[9]</sup>. Penyensoran data tahan hidup dapat disebabkan oleh beberapa hal, antara lain:

- Data hilang. Data tidak dapat diperoleh karena individu pengamatan hilang sebelum batas waktu pengamatan berakhir.
- Data keluar (*withdrawals*). Data tidak dapat diperoleh karena individu keluar dari pengamatan meskipun waktu pengamatan belum berakhir
- Waktu pengamatan berakhir. Kegagalan mungkin saja terjadi sebelum batas waktu pengamatan berakhir.

## 2.6. Metode Maksimum Likelihood (*Maximum Likelihood Estimator/MLE*)

Langkah-langkah untuk menentukan estimator maksimum likelihood dari  $\theta$  menurut Widiharih, 2003 adalah:

- Tentukan fungsi likelihood  $L(\theta_1, \theta_2, \dots, \theta_k | \tilde{X}) = \prod_{i=1}^n f(x_i | \theta_1, \theta_2, \dots, \theta_k)$ .
- Bentuk log likelihood  $l = \log L(\theta_1, \theta_2, \dots, \theta_k | \tilde{X})$
- Tentukan turunan dari  $l$  terhadap  $\theta_1, \theta_2, \dots, \theta_k$ ,  $\frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \dots, \theta_k | \tilde{X})$
- Bentuk persamaan likelihood dan selesaikan  $\frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \dots, \theta_k | \tilde{X}) = 0$

## 2.7. Kaplan-Meier Estimator

Metode Kaplan Meier digunakan untuk menaksir fungsi survival dan fungsi hazard. Metode ini disebut juga metode nonparametrik karena tidak membutuhkan asumsi distribusi dari waktu survival<sup>[4]</sup>.

### a. Taksiran Fungsi Survival

Misalkan terdapat  $n$  individu dengan waktu survival yaitu,  $t_1, t_2, \dots, t_n$ . Beberapa pengamatan ini tersensor, jika terdapat  $r$  waktu *failure* diantara  $n$  individu, dimana  $r \leq n$ , maka waktu failure ke- $j$  ditunjukkan sebagai  $t_{(j)}$ , untuk  $j = 1, 2, \dots, r$  dan  $r$  waktu *failure* adalah  $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ . Estimasi fungsi survival pada waktu ke- $k$  adalah

$$\hat{S}(t) = \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right) \quad (16)$$

dengan:  $n_j$  : jumlah pasien yang beresiko gagal pada  $t_j$

$d_j$  : Jumlah individu yang gagal pada waktu  $t_1, t_2, \dots, t_k$

### b. Taksiran Fungsi Hazard

Dalam menaksir fungsi hazard dari waktu survival menggunakan rasio jumlah *failure* terhadap jumlah individu yang berada pada resiko *failure*. Apabila  $d_j$  merupakan jumlah individu *failure* pada  $t_{(j)}$ , waktu survival ke- $j$  dan  $n_j$  adalah individu yang beresiko *failure* pada waktu  $t_{(j)}$ , maka estimasi fungsi hazard adalah

$$\hat{h}(t) = \frac{d_j}{n_j - \tau_j} \quad (17)$$

## 3. Pembahasan

Untuk memodelkan data survival dengan variabel penjelas yang mempengaruhi fungsi hazard adalah model hazard proporsional yang diusulkan oleh Cox dan juga dikenal sebagai regresi Cox<sup>[5]</sup>. Kelebihan Cox adalah tidak harus memiliki fungsi dari distribusi parametrik. Asumsi pemodelan hanya memvalidasi asumsi bahwa fungsi hazard harus proporsional setiap waktu. Asumsi proporsional pada model dapat diketahui melalui plot  $^{\circ}\log [-^{\circ}\log S(t)]$  terhadap waktu survival ( $t$ ) untuk setiap kategori yang ada dalam  $p$  variabel penjelas yang membentuk pola yang sejajar pada level yang berbeda-beda.

### 3.1. Model Hazard Proporsional

Misal hazard tergantung pada nilai-nilai  $x_1, x_2, \dots, x_p$  dari  $p$  variabel penjelas,  $X_1, X_2, \dots, X_p$ . Nilai-nilai dari variabel penjelas dalam model hazard proporsional dinyatakan dalam bentuk vektor  $\mathbf{x}$ , sehingga  $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ . Misalkan  $h_0(t)$  adalah fungsi hazard untuk individu yang semua variabel penjelas vektor  $\mathbf{x}$  mempunyai nilai nol, maka fungsi  $h_0(t)$  disebut *baseline hazard function*.

Model hazard proporsional umum adalah sebagai berikut

$$h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) h_0(t) \quad (18)$$

dapat pula dinyatakan dalam bentuk persamaan log,

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (19)$$

dengan  $h_i(t)$  : Fungsi hazard untuk individu ke- $i$

$h_0(t)$  : Fungsi hazard baseline

$\beta_1, \beta_2, \dots, \beta_p$  : Koefisien regresi

$x_{i1}, x_{i2}, \dots, x_{ip}$  : Nilai variabel untuk individu ke- $i$

### 3.1.3. Taksiran Parameter Model

Untuk menentukan model terbaik diperlukan taksiran koefisien variabel  $X_1, X_2, \dots, X_p$  yaitu  $\beta_1, \beta_2, \dots, \beta_p$ . Koefisien  $\beta$  dalam model hazard proporsional dapat ditaksir menggunakan metode Maksimum Likelihood. Apabila terdapat  $n$  individu, di antaranya terdapat  $r$  individu yang tidak tersensor dan  $n - r$  individu yang tersensor maka urutan waktu  $r$  waktu kegagalan dinotasikan oleh  $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ , sehingga  $t_{(j)}$  adalah urutan waktu kegagalan ke- $j$ . Menurut Cox (1972) fungsi likelihood untuk model hazard proporsional adalah

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta x_{(j)})}{\sum_{l \in R(t_j)} \exp(\beta x_{(j)})} \quad (20)$$

$x_{(j)}$  adalah vektor variabel dari individu yang gagal pada saat ke- $j$  dengan waktu  $t_j$ . Notasi  $R(t_j)$  adalah seluruh individu yang memiliki resiko gagal pada waktu ke- $j$ . Jika terdapat  $n$  waktu survival yang diobservasi, dinotasikan oleh  $t_1, t_2, \dots, t_n$  dan  $\delta_i$  adalah *value indicator* maka fungsi likelihoodnya dinyatakan dalam fungsi parsial likelihood sebagai berikut

$$L(\beta) = \prod_{j=1}^n \left[ \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_j)} \exp(\beta' x_{(j)})} \right]^{\delta_i} \quad (21)$$

dengan  $\delta_i = \begin{cases} 0, & \text{individu yang tersensor} \\ 1, & \text{individu tidak tersensor} \end{cases}$

Fungsi log likelihood yang bersesuaian adalah

$$\log L(\beta) = \sum_{j=1}^n \delta_i \{ \beta' x_{(j)} - \log \sum_{l \in R(t_j)} \exp^{\beta' x_l} \} \quad (22)$$

Estimasi koefisien  $\beta$  diselesaikan dengan metode numerik melalui penyelesaian iterasi Newton Raphson. Taksiran  $\beta$  pada iterasi ke  $(s+1)$ , yaitu

$$((\hat{\beta})_{s+1})_{px1} = ((\hat{\beta})_s)_{px1} + (I^{-1}(\hat{\beta})_s)_{pxp} (u(\hat{\beta})_s)_{px1} \quad (23)$$

dengan  $s = 0, 1, 2, \dots$

$(u(\hat{\beta})_s)_{px1} =$  vektor skor efisien berukuran  $px1$

$(I^{-1}(\hat{\beta})_s)_{pxp} =$  invers matriks informasi yang diamati berukuran  $pxp$

### 3.1.4. Taksiran Fungsi Survival dan Fungsi Hazard

Pada model hazard proporsional terdapat  $p$  variabel  $X_1, X_2, \dots, X_p$  dan taksiran koefisien dari variabel tersebut adalah  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  maka taksiran fungsi hazard ke- $i$  adalah

$$\hat{h}_i(t) = \exp^{\hat{\beta}' x_i} \hat{h}_0(t) \quad (24)$$

Dengan persamaan tersebut, fungsi hazard seorang individu dapat ditaksir

$$\hat{h}_0(t) = 1 - \xi_j \quad (25)$$

dengan,  $\xi_j = \left( 1 - \frac{\exp(\hat{\beta}' x_{(j)})}{\sum_{l \in R(t_j)} \exp(\hat{\beta}' x_l)} \right)^{\exp(-\hat{\beta}' x_{(j)})}$

Taksiran *baseline survivor function* dapat dihitung dengan

$$\hat{S}_0(t) = \prod_{j=1}^k \xi_j \quad \text{dimana } k = 1, 2, \dots, r \quad (26)$$

Nilai taksiran dari *baseline survivor function* adalah nol untuk  $t > t_{(r)}$ . Nilai taksiran dari *cumulative baseline hazard function* adalah

$$\hat{H}_0(t) = -\log \hat{S}_0(t) = -\sum_{j=1}^k \log \hat{\xi}_j \quad \text{dimana } k = 1, 2, \dots, r \quad (27)$$

### 3.2. Pengujian Parameter

#### 3.2.1. Pengujian Secara Serentak

- Hipotesis :  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$   
 $H_1 : \text{minimal ada satu dari } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$
- Taraf Signifikansi :  $\alpha = 0.05$
- Statistik uji :  $X_{LR}^2 = -2 \ln \left( \frac{l_0}{l_1} \right) = 2 (\ln l_1 - \ln l_0) = 2 (L_1 - L_0)$  (28)
- Kriteria uji : Tolak  $H_0$  jika  $X_{LR}^2 > \chi_{p; \alpha}^2$  atau sig  $< \alpha$   
 dimana  $p$  : jumlah  $\beta$   
 $L_1$  : Nilai  $-2 \log \tilde{L}$  dengan variabel  
 $L_0$  : Nilai  $-2 \log \tilde{L}$  tanpa variabel

#### 3.2.2. Pengujian Secara Parsial

- Hipotesis :  $H_0 : \beta_j = 0, \text{ untuk suatu } j \text{ dengan } j = 1, 2, \dots, p$   
 $H_1 : \beta_j \neq 0, \text{ untuk suatu } j \text{ dengan } j = 1, 2, \dots, p$
- Taraf Signifikansi :  $\alpha = 0.05$
- Statistik uji :  $X_W^2 = \left[ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2$  (29)
- Kriteria uji : Tolak  $H_0$  jika  $X_W^2 > \chi_{1; \alpha}^2$  atau sig  $< \alpha$

### 3.3. Aplikasi Metode Cox Proportional Hazard Model

#### 3.3.1. Sumber Data

Populasi dalam penelitian ini adalah semua penderita kanker leher rahim yang pernah menjalani perawatan di rumah sakit Dr. Cipto Mangunkusumo Jakarta periode Januari 1997 sampai dengan September 2002, sedangkan sampel yang diambil adalah penderita kanker leher rahim yang bertempat tinggal di Jakarta 80 orang. Data diperoleh dari data rekam medik rumah sakit. Sumber: Sirait, A. M., et al., Ketahanan Hidup Penderita Kanker Serviks di Rumah Sakit Dr. Cipto Mangunkusumo Jakarta, *Indones J. Obstet Gynecol*, 1997, Vol. 21, No. 3: 182-190.

#### 3.3.2. Identifikasi Variabel

Data-data yang diperlukan dalam penelitian.

- Variabel dependen berupa data waktu survival ( $t$ ) penderita kanker leher rahim dengan satuan bulan.
- Variabel independen yang digunakan dalam penelitian ini berupa data faktor-faktor yang diduga mempengaruhi ketahanan hidup penderita kanker leher rahim. meliputi
  - a. usia (U), dengan 3 kategori: 1 =  $20 \leq \text{usia} < 40$ , 2 =  $40 \leq \text{usia} < 60$ , 3 =  $60 \leq \text{usia} < 80$
  - b. stadium (S), dengan 4 kategori: 1 = stadium I, IA, IB, 2 = stadium II, IIA, IIB, 3 = stadium III, IIIA, IIIB, 4 = stadium IV, IVA, IVB
  - c. pengobatan awal (Pg), dengan 4 kategori: 0 = operasi, 1 = radioterapi, 2 = kemoterapi, 3 = imunoterapi, 4 = kombinasi 4 pengobatan
  - d. penyakit lain (Pl), dengan 2 kategori : 0 = tidak, 1 = ya
  - e. pendidikan terakhir (Pd), dengan 4 kategori: 1 = SD, 2 = SMP, 3 = SMA, 4 = Perguruan Tinggi
  - f. anemia (A), dengan 2 kategori: 0 = tidak, 1 = ya

Pemilihan variabel usia, stadium, pengobatan awal dan penyakit lain berdasarkan penelitian yang dilakukan oleh Wulandari (2001). Sedangkan variabel pendidikan terakhir

berdasarkan penelitian Sirait (1997). Variabel anemia, menurut Azis, dkk merupakan salah satu faktor yang mempengaruhi ketahanan hidup penderita kanker leher rahim.

### 3.4. Pemodelan Regresi Cox

#### 3.4.1. Model Cox Proportional Hazard Model

Dengan software SPSS, diperoleh model regresi Cox untuk responden ke-*i* sebagai berikut:

$$h_i(t) = \exp(-2.452U1_i - 1.946U2_i - 16.995S1_i - 2.96S2_i - 2.333S3_i + 1.835Pg1_i + 0.57Pg2_i + 1.46Pg3_i - 0.33Pg4_i - 0.49Pl_i - 0.578Pd1_i - 1.52Pd2_i - 1.78Pd3_i - 1.58A_i) h_0(t)$$

dilakukan pengujian serentak dan parsial pada model awal dengan perhitungan statistik

Hipotesis :  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$   
 $H_1 : \text{minimal ada } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$

Taraf Signifikansi :  $\alpha = 0.05$

Statistik Uji :  $\chi^2_{LR} = -2 \ln \left( \frac{l_0}{l_1} \right)$   
 $= 2 (\ln l_1 - \ln l_0)$   
 $= 2 (L_1 - L_0)$   
 $= 287.604 - 355$   
 $= 68.282$

Kriteria Uji : Tolak  $H_0$  jika  $\chi^2_{LR} > \chi^2_{p; \alpha}$   
 atau tolak  $H_0$  jika  $\text{Sig} < \alpha (0.05)$

Keputusan : Dari tabel distribusi  $\chi^2$  diperoleh nilai  $\chi^2_{14;0.05} = 23.6848$ .  
 Karena  $\chi^2_{LR} = 68.282 > \chi^2_{p; \alpha} = \chi^2_{14;0.05} = 23.6848$   
 atau  $\text{Sig} = 0.000 < \alpha = 0.05$ , maka  $H_0$  ditolak.

Kesimpulan : model secara keseluruhan adalah signifikan.

Selanjutnya dilakukan uji parsial terhadap model awal.

Hipotesis :  $H_0 : \beta_j = 0$ , untuk suatu  $j$  dengan  $j = 1, 2, \dots, p$   
 $H_1 : \beta_j \neq 0$ , untuk suatu  $j$  dengan  $j = 1, 2, \dots, p$

Taraf Signifikansi :  $\alpha = 0.05$

Statistik Uji :  $X^2_W = \left[ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2 = -2.452^2 / 0.767^2 = 10.228$

Kriteria Uji : tolak  $H_0$  jika  $X^2_W = \left[ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2 > \chi^2_{1;0.05} = 3.841$   
 atau  $\text{Sig} < \alpha (0.05)$

Berdasarkan output SPSS 16 diperoleh tabel 1.

**Tabel 1.** Uji Parsial Model

Variabel	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	$X^2_W$	Sig	Keputusan
20 ≤ Usia < 40 (U1)	-2.452	0.767	10.228	<b>0.001</b>	<b>Tolak H<sub>0</sub></b>
40 ≤ Usia < 60 (U2)	-1.946	0.474	16.855	<b>0.000</b>	<b>Tolak H<sub>0</sub></b>
Stadium 1 (S1)	-6.996	409.282	0.0017	0.967	Terima H <sub>0</sub>
Stadium 2 (S2)	-2.96	0.724	16.715	<b>0.000</b>	<b>Tolak H<sub>0</sub></b>
Stadium 3 (S3)	2.33	0.706	10.892	<b>0.001</b>	<b>Tolak H<sub>0</sub></b>
Operasi (Pg1)	1.83	1.017	3.238	0.071	Terima H <sub>0</sub>
Radioterapi (Pg2)	0.57	0.557	1.047	0.307	Terima H <sub>0</sub>
Kemoterapi (Pg3)	1.46	0.853	2.929	0.088	Terima H <sub>0</sub>



Imunoterapi (Pg4)	- 0.33	0.437	0.570	0.451	Terima H <sub>0</sub>
Penyakit Lain (PI)	- 0.49	0.421	1.355	0.246	Terima H <sub>0</sub>
SD (Pd1)	- 0.578	1.146	0.254	0.614	Terima H <sub>0</sub>
SMP (Pd2)	- 1.52	1.190	1.631	0.203	Terima H <sub>0</sub>
SMA (Pd3)	- 1.78	1.234	2.081	0.150	Terima H <sub>0</sub>
Anemia (A)	- 1.58	0.614	6.622	<b>0.010</b>	<b>Tolak H<sub>0</sub></b>

Pada tabel 1 diketahui bahwa variabel pengobatan awal, penyakit lain dan pendidikan terakhir tidak signifikan dalam model awal.

Pengujian serentak pada model terbaik

Hipotesis : H<sub>0</sub> : β<sub>1</sub> = β<sub>2</sub> = β<sub>3</sub> = β<sub>4</sub> = 0  
 H<sub>1</sub> : minimal ada β<sub>j</sub> ≠ 0, dengan j = 1,2,3,4

Statistik Uji :  $X_{LR}^2 = -2 \ln \left( \frac{l_0}{l_1} \right)$   
 $= 2 (\ln l_1 - \ln l_0)$   
 $= 2 (L_1 - L_0)$   
 $= 355.886 - 300.19$   
 $= 55.689$

Kriteria Uji : Tolak H<sub>0</sub> jika  $X_{LR}^2 > \chi^2_{p; \alpha}$   
 atau tolak H<sub>0</sub> jika Sig < α (0.05)

Keputusan : Dari tabel distribusi  $\chi^2$  diperoleh nilai  $\chi^2_{4;0.05} = 9.48773$ .  
 Karena  $X_{LR}^2 = 55.689 > \chi^2_{p; \alpha} = \chi^2_{4;0.05} = 9.48773$  atau Sig = 0.000 < α = 0.05 maka H<sub>0</sub> ditolak.

Kesimpulan bahwa pemodelan secara keseluruhan adalah signifikan.

Uji Parsial Terhadap Model Terbaik

Hipotesis : H<sub>0</sub> : β<sub>j</sub> = 0, j = 1,2,3,4  
 H<sub>1</sub> : β<sub>j</sub> ≠ 0, j = 1,2,3,4

Statistik Uji :  $X_W^2 = \left\{ \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right\}^2$

Kriteria Uji : tolak H<sub>0</sub> jika  $X_W^2 > \chi^2_{1;0.05} = 3.841$  atau Sig < α (0.05)

pemodelan dengan variabel usia dan stadium dapat disajikan dalam tabel 2 berikut ini

**Tabel 2.** Uji Parsial Model Terbaik

Variabel	$\hat{\beta}_j$	SE( $\hat{\beta}_j$ )	Exp( $\hat{\beta}_j$ )	$X_W^2$	Sig	Keputusan
20 ≤ Usia < 40 (U1)	-1.848	0.679	0.158	7.407	0.007	<b>Tolak H<sub>0</sub></b>
40 ≤ Usia < 60 (U2)	-1.584	0.391	0.205	16.412	0.000	<b>Tolak H<sub>0</sub></b>
Stadium 2 (S2)	-3.255	0.625	0.039	27.123	0.000	<b>Tolak H<sub>0</sub></b>
Stadium 3 (S3)	-2.108	0.580	0.122	13.21	0.000	<b>Tolak H<sub>0</sub></b>

Dari perhitungan pada tabel 2 diketahui bahwa faktor usia dan stadium klinik signifikan dalam model terbaik. Dengan demikian dapat diartikan bahwa faktor yang mempengaruhi ketahanan hidup responden adalah usia dan stadium. Interpretasi model terbaik adalah responden dengan usia 20 hingga 39 tahun memiliki peluang meninggal 0,158 kalinya penderita kanker leher rahim usia lain, sedangkan responden dengan usia 40 hingga 59 tahun memiliki peluang meninggal 0,205 kalinya penderita kanker leher rahim usia lain, untuk responden dengan stadium II memiliki peluang meninggal 0,039 kalinya

penderita kanker leher rahim stadium lain dan responden dengan stadium III memiliki peluang meninggal 0,122 kalinya penderita kanker leher rahim stadium lain.

#### 4. Kesimpulan

Berdasarkan fungsi survival, responden mempunyai harapan hidup sebesar 43,8% untuk hidup selama 5 tahun sejak pertama kali menjalani pengobatan. Pemodelan regresi untuk penderita kanker leher rahim di RSUD Dr. Cipto Mangunkusumo Jakarta adalah usia dan stadium dengan model sebagai berikut :

$$h_i(t) = \exp(-1.848U1_i - 1.584U2_i - 3.255S2_i - 2.108S3_i) h_0(t)$$

Berdasarkan model tersebut dapat diketahui bahwa responden dengan usia 20 hingga 39 tahun memiliki peluang meninggal 0,158 kalinya penderita kanker leher rahim usia lain, sedangkan responden dengan usia 40 hingga 59 tahun memiliki peluang meninggal 0,205 kalinya penderita kanker leher rahim usia lain, untuk responden dengan stadium II memiliki peluang meninggal 0,039 kalinya penderita kanker leher rahim stadium lain dan responden dengan stadium III memiliki peluang meninggal 0,1215 kalinya penderita kanker leher rahim stadium lain.

#### DAFTAR PUSTAKA

1. Azis, dkk., *Pengaruh Anemia Terhadap Penderita Kanker Leher Rahim*, RS Dr. Cipto Mangunkusumo Jakarta, Jakarta, 2007.
2. Bain, L.J., and Engelhardt, M., *Introduction to Probability and Mathematical Statistics*, Duxbury press, California, 1992.
3. Blischke, W.R., and Murthy, P.D.N., *Reliability: Modelling, Prediction, and Optimization*, John Wiley & Sons, Inc., Canada, 2000.
4. Collet, D., *Modelling Survival Data in Medical Research*, Chapman and Hall, London, 1994.
5. Cox, D. R., Regression Models and Life Tables (with discussion), *Journal of The Royal Statistical Society B*, 1972, 34:187-220.
6. Cox and Oakes, D., *Analysis of Survival Data*, Chapman and Hall, London, 1984.
7. Elsayed, E.A., *Reliability Engineering*, Addison Wesley Longman, New York, 1996.
8. Lawless, J.F., *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc., Canada, 1982.
9. Lee, E.T., *Statistical Methods for Survival Data Analysis*, John Willey and Sons, Inc., New York, 1992.
10. Miles, M.B. and Huberman, A.M., *Analisis Data Kualitatif*, diterjemahkan oleh Tjetjep Rohendi Rohidi, Universitas Indonesia Press, Jakarta, 1992.
11. Miller, G.R., *Survival Analysis*, Statistic Department: California Stanford University, 1998.
12. Sirait, A.M., et al., Ketahanan Hidup Penderita Kanker Serviks di Rumah Sakit Dr. Cipto Mangunkusumo Jakarta, *Indones J. Obstet Gynecol*, 1997, Vol. 21, No. 3: 182-190.
13. Widiharih, T., dan Suparti, *Buku Ajar Statistika Matematika II*, Laboratorium Statistika Jurusan Matematika FMIPA UNDIP, Semarang, 2003.
14. Wulandari, *Analisis Survival Terhadap Pasien Kanker Leher Rahim*, RS Dr. Cipto Mangunkusumo Jakarta, Jakarta, 2001.