

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Jenis Sumber Data**

Data yang digunakan dalam penelitian ini merupakan data simulasi berdasarkan studi yang dilakukan oleh Henderson (2002). Studi tersebut berfokus pada data *survival* pasien yang menderita penyakit AML di Northwest England selama periode 1982 hingga 1998. Pada rentang waktu tersebut, sebanyak 100 individu penderita AML diamati untuk tujuan penelitian. Data ini mengandung informasi yang sangat berharga tentang berbagai faktor yang berpotensi memengaruhi waktu tahan hidup atau *survival* rate pasien.

Data yang dikumpulkan tidak hanya mencakup variabel utama seperti waktu bertahan hidup pasien, tetapi juga berbagai faktor tambahan yang dapat memberikan wawasan lebih mendalam tentang penyebab variasi dalam *survival* rate. Faktor-faktor tersebut meliputi umur pasien, jenis kelamin, tingkat ketidakmampuan daerah, dan jumlah sel darah putih.

#### **3.2 Metode Analisis Data**

Metode analisis data yang digunakan dalam penelitian ini terdiri dari beberapa pendekatan yang berbeda, masing-masing memiliki tujuan dan fungsi yang spesifik dalam memberikan gambaran yang lebih lengkap dan mendalam mengenai data yang dikumpulkan. Pendekatan-pendekatan ini mencakup analisis deskriptif dan analisis regresi *Cox proportional hazard* yang berdistribusi Gompertz, dengan memanfaatkan metode Estimasi Kemungkinan Maksimum dan metode Newton-Raphson.

#### **3.3 Variabel Data**

Data yang digunakan merupakan data simulasi yang terdiri dari umur, jenis kelamin, status merokok, jumlah sel darah putih (wbc), dan tingkat kemakmuran daerah (tpi). Sebagai pertimbangan efek acak spasial terhadap waktu tahan hidup, akan ditambahkan lokasi tiap individu berupa koordinat.

Adapun definisi dari variable yang digunakan dalam penelitian adalah:

1. Waktu tahan hidup ( $t$ )  
Waktu sampai terjadinya waktu kejadian. Waktu yang digunakan pada penelitian yaitu lamanya penderita (dalam hari) menderita leukemia hingga meninggal.
2. Indikator sensor (cens)  
Status sensor individu (0 jika masih hidup dan 1 jika mengalami kejadian atau meninggal).
3. Koordinat tempat tinggal ( $coorx, coory$ )  
Lokasi titik tiap individu di mana individu tersebut tinggal.
4. Umur (age) ( $X_1$ )  
Umur saat individu menderita penyakit leukemia.
5. Jenis Kelamin (sex) ( $X_2$ )  
Penderita leukemia berjenis kelamin laki-laki dan perempuan dalam bentuk kategori (1 jika penderita berjenis kelamin laki-laki dan 0 jika penderita berjenis kelamin Perempuan)
6. Jumlah sel darah putih (wbc) ( $X_3$ )  
Jumlah sel darah putih yang dimiliki individu dengan 1 unit =  $5000/\mu L$
7. Tingkat ketidakkakmuran daerah ( $tpi$ ) ( $X_4$ )  
Nilai  $tpi$  yang semakin kecil menunjukkan menunjukkan daerah yang lebih makmur.

Tabel 3.1 Variabel Kovariat Data AML

Variabel	Keterangan	Jenis	Kategori
$Y$	Waktu ketahanan hidup (dalam hari)	Kontinu	-
$X_1$	Usia	Kontinu	-
$X_2$	Jenis Kelamin	Kategorikal	1 =perempuan; 0 = laki-laki
$X_3$	Jumlah sel darah putih	Kontinu	-

$X_4$	Tingkat Ketidakmakmuran Daerah	Kontinu	-
-------	--------------------------------------	---------	---

### 3.4 Model Faktor Acak Spasial

Banerjee, dkk. mengembangkan model *survival* dengan melibatkan ketergantungan efek acak pada data spasial. Efek acak  $W$  mewakili korelasi antar observasi pada lokasi-lokasi tertentu, sehingga efek acak spasial dapat dirumuskan sebagai berikut.

$$h(t_{ij}; x_{ij}) = h_0(t) \exp(\beta^T x_{ij} + W_i)$$

Distribusi untuk *frailty* dalam model geostatistik adalah:

$$W \sim N_n(0, Z(\theta)) \quad (3.1)$$

dengan:

$$Z(\theta) = \sigma^2 \exp(-\phi d_{ij}); \sigma^2 > 0, \phi > 0, d > 0 \quad (3.2)$$

### 3.5 Model *Survival* Spasial

Misalkan  $T$  merupakan variable acak waktu ketahanan hidup. Setiap sampel individu diasumsikan berkaitan dengan waktu kegagalan. Sebagian besar data *survival* merupakan observasi pada individu yang terdiri dari kovariat yang saling berkaitan dengan individu.

Fungsi *hazard* kumulatifnya adalah:

$$\begin{aligned} H(t_{ij}; x_{ij}) &= \int_0^{t_{ij}} h(u_{ij}; x_{ij}) du_{ij} \\ &= \int_0^{t_{ij}} e^{au_{ij}} \exp(\beta^T x_{ij} + W_i) du_{ij} \\ &= \left( \frac{1}{a} e^{au_{ij}} \exp(\beta^T x_{ij} + W_i) \right)_{u_{ij}=0}^{t_{ij}} \\ &= \left( \frac{1}{a} \exp(\beta^T x_{ij} + W_i) \right) (e^{at_{ij}} - 1) \end{aligned}$$

Melihat hubungan antara fungsi *survival* dan fungsi *hazard* kumulatif, diperoleh:

$$\begin{aligned} S(t_{ij}; x_{ij}) &= \exp(-H(t_{ij}; x_{ij})) \\ &= \exp\left(-\left(\frac{1}{a} \exp(\beta^T x_{ij} + W_i)\right) (1 - e^{at_{ij}})\right) \end{aligned}$$

dengan fungsi densitas diperoleh sebagai berikut.

$$f(t_{ij}; x_{ij}) = h(t_{ij}; x_{ij})S(t_{ij}; x_{ij})$$

$$f(t_{ij}; x_{ij}) = e^{at_{ij}} \exp(\beta^T x_{ij} + W_i) \cdot \exp\left(\left(\frac{1}{a} \exp(\beta^T x_{ij} + W_i)\right)(1 - e^{at_{ij}})\right)$$

Jadi, model *survival* spasial yang diperoleh untuk model proporsional *hazard* parametrik adalah:

$$S(t_{ij}; x_{ij}) = \exp\left(\left(\frac{1}{a} \exp(\beta^T x_{ij} + W_i)\right)(1 - e^{at_{ij}})\right)$$

Dengan  $t$  adalah waktu ketahanan hidup,  $\beta_1, \beta_2, \dots, \beta_p$  adalah parameter kovariat, dan  $h_0(t)$  adalah fungsi *baseline hazard*.  $S(t_{ij}; x_{ij})$  dapat diartikan sebagai peluang ketahanan hidup individu setelah waktu  $t$  dengan variable kovariat  $X_1, X_2, \dots, X_p$ .

Model parametrik dapat dilakukan melalui hasil fungsi *survival* likelihood. Jika dimiliki sampel acak dari pasangan  $(T_{ij}, \delta_{ij}), i = 1, 2, \dots, l, j = 1, 2, \dots, n_i$  maka fungsi likelihoodnya adalah:

$$L = \prod_{i=1}^l \prod_{j=1}^{n_i} [f(t_{ij}; x_{ij})]^{\delta_{ij}} [S(t_{ij}; x_{ij})]^{1-\delta_{ij}}$$

dengan  $\delta_{ij}$  mengindikasikan kehidupan individu ke- $j$  wilayah  $i$  masih hidup ( $\delta_{ij} = 1$ ) atau tidak ( $\delta_{ij} = 0$ ). Dengan demikian, fungsi *survival* likelihoodnya adalah:

$$L = \prod_{i=1}^l \prod_{j=1}^{n_i} [e^{at_{ij}} \exp(\beta^T x_{ij} + W_i)]^{\delta_{ij}} \exp\left(\left(\frac{1}{a} \exp(\beta^T x_{ij} + W_i)\right)(1 - e^{at_{ij}})\right) \quad (3.3)$$

Berikutnya adalah menentukan fungsi log natural kemungkinan. Hal tersebut dikarenakan perhitungannya yang lebih sederhana dan hasil yang tidak berbeda dengan fungsi kemungkinan. Berdasarkan persamaan (3.3), diperoleh:

$$l = \sum_{i,j} \delta_{ij} (at_{ij} + \beta^T x_{ij} + W_i) + \left(\frac{1}{a} \exp(\beta^T x_{ij} + W_i)(1 - e^{at_{ij}})\right)$$

Misalkan  $\mu_{ij} = \beta^T x_{ij} + W_i$ , maka bentuk sederhana dari log kemungkinan untuk model *Cox proportional hazard* parametrik berdistribusi Gompertz dapat dinyatakan sebagai berikut.

$$l = \sum_{i,j} \delta_{ij} (at_{ij} + \mu_{ij}) + \left( \frac{1}{a} e^{\mu_{ij}} (1 - e^{at_{ij}}) \right) \quad (3.4)$$

Langkah-langkah yang akan dilakukan dalam penelitian adalah:

1. Menentukan fungsi *hazard* berdistribusi Gompertz
2. Memodelkan fungsi *hazard* dengan *frailty*  $W_i$  ke dalam model *Cox Proportional Hazard*
3. Menentukan fungsi likelihood dari fungsi yang dibangun berdasarkan model *proportional hazard* dengan *frailty*
4. Mengestimasi parameter dengan metode Newton-Raphson.
5. Menerapkan model *proportional hazard* berdistribusi Gompertz dengan *frailty* dalam bentuk model *survival*  $S(t_{ij}; x_{ij})$

### 3.6 Teknik Analisis Data

Software R akan digunakan untuk melakukan membangkitkan data, analisis statistika secara deskriptif, melakukan estimasi parameter, dan melakukan pengujian parameter. Dalam penelitian tersebut, teknik analisis data terdiri dari:

1. Membuat data simulasi  
Tahapan awal penelitian melibatkan data sekunder. Data sekunder digunakan sebagai dasar untuk melakukan analisis statistik dan pemodelan.
2. Analisis deskriptif terhadap variabel bebas serta variabel tak bebas  
Analisis deskriptif dilakukan terhadap variabel-variabel yang dimiliki. Hal tersebut mencakup mean, nilai maksimum, nilai minimum, dan visualisasi data melalui grafik.
3. Uji distribusi tertentu dari waktu dengan menggunakan *Anderson-Darling*  
Langkah selanjutnya adalah menguji distribusi waktu *survival* data. Uji *Anderson-Darling* adalah salah satu uji statistik yang digunakan untuk menguji apakah data waktu ketahanan hidup berdistribusi Gompertz.
4. Memasukkan efek acak spasial  
Setelah mengetahui data berdistribusi Gompertz, efek acak spasial dapat

dimasukkan ke dalam model *survival*. Efek tersebut memungkinkan penanganan variabilitas terkait dengan lokasi geografis atau elemen-elemen lainnya yang tidak dapat diketahui.

5. Estimasi parameter variable bebas menggunakan MLE dan Newton-Raphson  
Tahap melibatkan variabel-variabel bebas.
6. Uji signifikansi parameter menggunakan *Z Wald*
7. Pemodelan *Cox Proportional Hazard* dan *survival* spasial dilakukan
8. Menghitung peluang hidup individu berdasarkan model *Cox Proportional Hazard* dan *survival* spasial
9. Kesimpulan

### 3.7 Diagram Alir

