# BAB II

# TINJAUAN PUSTAKA

## 2.1 Pendahuluan

Pada bab ini akan dipaparkan teori-teori yang menjadi dasar dan landasan dalam pembuatan skripsi ini sehingga membantu mempermudah pembahasan bab selanjutnya dan pembahasan utama dalam skripsi ini. Teori tersebut meliputi analisis *survival*, data tersensor, estimator Kaplan-Meier, pemeriksaan pengamatan *outlier* dan pengamatan yang berpengaruh terhadap estimator fungsinya, perubahan nilai kemungkinan (*likelihood displacement*), dan statistik rasio kemungkinan (*likelihood ratio statistic*).

## 2.2 Kanker Payudara

Kanker payudara adalah tumor ganas yang tumbuh di dalam jaringan payudara. Kanker bisa mulai tumbuh di dalam kelenjar susu, jaringan lemak, maupun jaringan ikat pada payudara (Sastrosudarmo, 2011).

Tubuh manusia terdiri dari sel-sel yang selalu tumbuh, kadang-kadang pertumbuhan sel tersebut tidak terkontrol dan membentuk suatu gumpalan. Apabila pada satu tempat di tubuh manusia, salah satu contoh adalah jaringan payudara dimana seharusnya ketika ada sel yang rusak, sel tersebut akan mati dan digantikan oleh sel yang baru, tetapi jika pada proses ini terjadi kelainan dimana sel yang usang tadi tidak langsung mati tetapi membangun sel tambahan yang tidak sesuai dengan kebutuhan tubuh maka terjadilah pertumbuhan sel-sel yang berlebihan, dan membentuk suatu benjolan atau tumor di payudara. Tumor ini dapat bersifat jinak maupun ganas, tumor yang ganas inilah yang disebut dengan kanker, apabila berada di organ payudara maka disebut dengan kanker payudara (Pamungkas, 2011).

Kanker payudara juga dapat dijelaskan sebagai suatu kondisi dimana pertumbuhan sel yang ada di payudara telah kehilangan pengendalian dalam mekanisme normalnya sehingga mengalami pertumbuhan yang tidak normal, cepat dan tidak terkendali dari pada sel - sel kelenjar maupun salurannya.

Penyebab pasti kanker payudara sampai saat ini belum diketahui, namun sel kanker disebabkan oleh adanya genom abnormal yang terjadi karena adanya kerusakan gen yang mengatur pertumbuhan dan diferensiasi sel.

Ada beberapa faktor resiko yang bisa meningkatkan kemungkinan terjadinya kanker payudara. Beberapa diantaranya adalah sebagai berikut:

a. Usia, resiko kanker payudara semakin meningkat dengan bertambahnya umur (diatas 30 tahun).
b. Riwayat keluarga, wanita yang ibu atau saudara perempuannya pernah menderita kanker, memiliki resiko 3 kali lebih besar untuk mengalami kanker payudara.
c. Faktor hormon, hormon merupakan hal yang paling banyak berpengaruh terhadap kanker payudara, seperti mendapat haid pertama sebelum berusia 12 tahun, menopause setelah umur 55 tahun, tidak menikah atau tidak pernah melahirkan anak pertama setelah berusia 35 tahun serta pengguna pil KB lebih dari 3 tahun atau terapi hormon estrogen.
d. Faktor genetik, terdapat 2 varian gen BRCA1 dan BRCA2 yang merupakan suatu gen suspeptabilitas kanker payudara, jika salah satu wanita memiliki satu gen tersebut maka kemungkinan untuk menderita kanker payudara amatlah besar.

e. Ras, wanita kulit putih kemungkinan kecil menderita kanker payudara dibandingkan wanita Afrika - Amerika kulit hitam, karena wanita Afrika mempunyai tumor yang masa tumbuhnya lebih cepat dan berdampak kepada kematian karena kanker payudara.

f. Radiasi, pemaparan terhadap penyinaran (radiasi) terutama pada bagian dada, dan pernah menjalani terapi radiasi di bagian dada dimana pernah menderita kanker lain seperti limfoma secara signifikan mengalami peningkatan untuk terkena kanker payudara.

Dari seluruh penderita kanker payudara yang tergabung dalam Klub Ken Dedes Jakarta hanya penderita yang pernah mengalami operasi yang dianalisis dalam kajian ini. Datanya berukuran 44 amatan, dimana ada 19 penderita kanker payudara yang meninggal. *Survival time*-nya diukur dalam bulan sejak dilakukannya operasi.

## 2.3 Analisis *Survival*

Analisis *survival* dikembangkan pertama kali oleh astronom Inggris, yaitu Edmund Halley (1656-1742) (Armitage, 1973, Johnson & Johnson, 1980, Miller, 1981, Kuzma, 1984). Analisis ini menjadi salah satu alat penting dalam statistik vital dan ilmu aktuarial serta ilmu lainnya. *Survival* merupakan asal kata dari *to survive* yang berarti ketahanan atau kelangsungan hidup (Kleinbaum, 1996, Johnson & Johnson, 1980, Miller, 1981).

Analisis *survival* merupakan analisis data, dimana datanya adalah "lamanya waktu" yang dimulai dari suatu "titik asal" sampai terjadinya suatu "kejadian" atau titik akhir. Waktu dapat berupa tahun, bulan, hari, jam, atau bahkan menit yang diukur sejak pengamatan dimulai hingga muncul kejadian. Kajian ini disebut dengan analisis reliabilitas dalam ilmu teknik, analisis durasi dalam ilmu ekonomi dan analisis sejarah kejadian (*event history*) dalam ilmu sosial. Variabel responnya sering disebut sebagai waktu masa hidup (*survival time*), waktu kerusakan (*failure time*) atau waktu sampai terjadinya kejadian (*time-to-event*). Kejadian yang diamati dapat berupa kematian, insiden penyakit, kekambuhan, atau penyembuhan (Kleinbaum, 1996).

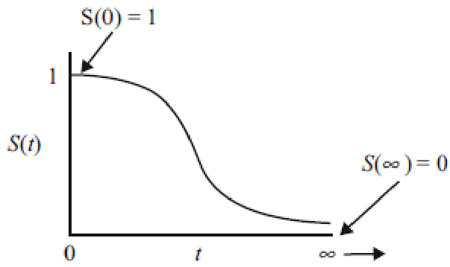
Adapun sifat khas dari data *survival*, yaitu: (1) pengamatannya non-negatif, (2) tidak berdistribusi normal, melainkan miring ke kanan (*right skewed*), (3) ada individu yang datanya tidak dapat ditentukan, dan (4) biasanya tidak semua individu masuk kedalam penelitian pada saat yang sama, ada pasien yang masuk belakangan.

Ada dua metode yang digunakan dalam analisis *survival* yaitu metode parametrik, dan nonparametrik. Metode parametrik diasumsikan bahwa data antar kejadian berdistribusi tertentu seperti eksponensial, gamma, weibull dan lain sebagainya. Sedangkan metode nonparametrik adalah metode yang tidak bergantung pada asumsi distribusi populasinya dan sering disebut dengan metode bebas distribusi (*distribution-free method*). Ada tiga metode nonparametrik yang sering digunakan dalam analisis *survival* yaitu metode Kaplan-Meier, *Life Table*, dan Nelson-Aalen.

Ciri khas dari analisis data *survival* adalah fungsi *survival* yang dinotasikan dengan $S(t)$ didefinisikan sebagai probabilitas dari suatu individu untuk bertahan setelah waktu yang ditetapkan. Secara teori, fungsi *survival* dapat digambarkan dengan kurva mulus dan memiliki karakteristik (Kleinbaum dan Klein, 2005):

1. Tidak naik, kurva cenderung menurun ketika *t* bertambah.
2. Untuk $t = 0, S(t) = S(0) = 1$.
3. Untuk $t = \infty, S(t) = S(\infty) = 0$.

Gambar 2.1 menunjukkan kurva fungsi *survival*.

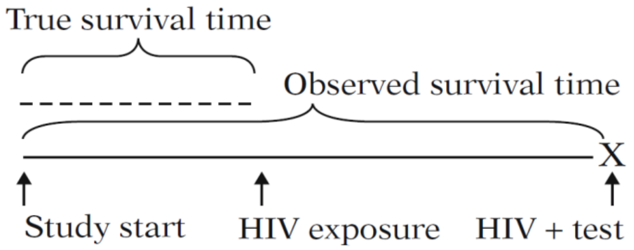



**Gambar 2.1** Kurva fungsi *survival*

## 2.4 Data Tersensor

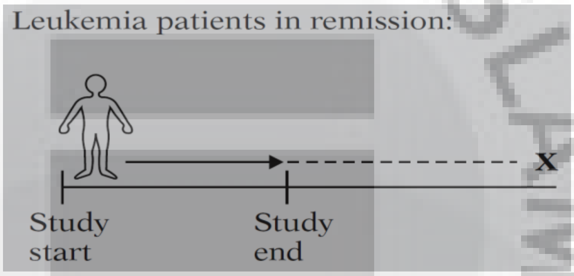
Pada analisis *survival* selalu terjadi data tersensor (*censored data*), yaitu ada informasi mengenai waktu *survival* individu tetapi tidak diketahui secara pasti berapa lama waktu *survival* (Kleinbaum, 1996). Penyebab terjadinya adalah hingga studi berakhir belum muncul kejadian yang diinginkan, hilang dari pengamatan, atau mengalami kejadian yang tidak berhubungan dengan substansi yang diteliti. Ada 3 jenis penyensoran, yaitu:

1. *Left-censored* (tersensor kiri), observasi dikatakan *left-cencored* jika objek yang diobservasi mengalami peristiwa di bawah waktu yang telah ditetapkan atau ketika masa observasi belum selesai. Misalnya, Penelitian tentang lamanya waktu sampai seseorang menjadi positif HIV, dimana kita akan mencatat kejadian tsb manakala dia dinyatakan positif ketika dilakukan test pertama kali. Akan tetapi kita tidak tahu secara persis kapan sebenarnya dia terjangkit virus tsb, sehingga waktu kejadiannya tidak diketahui. Hal ini menggambarkan sensor kiri, karena waktu survivalnya adalah sampai terjangkit HIV yg lebih pendek dibandingkan dengan waktu teramatinya, yang baru diketahui saat dilakukan test.



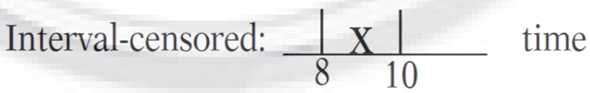

**Gambar 2.2** Ilustrasi Pengamatan Tersensor Kiri

2. *Right-censored* (tersensor kanan), obsevasi dikatakan *right-cencored* jika objek masih hidup atau masih beroperasi ketika masa observasi telah selesai. Misalnya, Pasien tersebut belum mengalami kejadian yg diteliti saat penelitian berakhir, Pasien tidak mengontak lagi (lost to follow-up) selama masa penelitian, Pasien keluar dari penelitian.




**Gambar 2.3** Ilustrasi Pengamatan Tersensor Kanan

3. *Interval-censored* (sensor interval), ketika objek mengalami peristiwa diantara interval waktu tertentu maka observasi dikatakan *interval-censored.* Misalnya, suatu individu mengalami kejadian antara umur 8 sampai 10 tahun (waktu persisnya tidak diketahui). Individu ini tersensor interval (yakni, 8 < $t$ < 10).

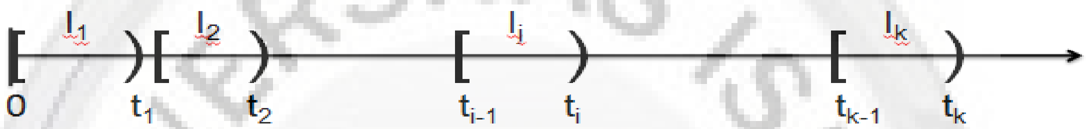



**Gambar 2.4** Ilustrasi Pengamatan Tersensor Interval

## 2.5 Estimator Kaplan-Meier

Telah diketahui bahwa salah satu tujuan dari analisis *survival* ialah mengestimasi dan menginterpretasi fungsi *survival*. Banyak metode yang digunakan untuk mengestimasi fungsi kesintasan, diantaranya *Nelson-Aalen estimator*, metode *life-table (acturial)*, metode Kaplan-Meier, dan lain-lain. Namun dalam penulisan skripsi ini metode yang digunakan adalah metode Kaplan-Meier yang merupakan statistik nonparametrik dengan data tersensor kanan, sehingga penggunaan metode Kaplan-Meier adalah yang paling baik.

Misalkan ada *n* individu dengan waktu pengamatan $x_1, x_2, ..., x_n$ dan ada *k* individu yang meninggal, dimana $k \leq n$. Waktu meninggalnya diurutkan $t_{(1)} < t_{(2)} < ... < t_{(k)}$. Misal data waktu survival kita buat menjadi *k* buah interval yakni $I_1$, ..., $I_k$




Dari interval tersebut dapat didefinisikan:

$r_j$ : banyaknya subjek yang masih hidup dalam penelitian pada saat $t_j$ pada interval $I_j$

$d_j$ : banyaknya subjek yang mati saat $t_j$ pada interval $I_j$

P(mati pada $[t_{(k)} - \Delta, t_{(k)}]$) = $d_j / r_j$ ... (2.1)

Dimana $\Delta$ adalah lebar selang waktu yg kecil.

P(bertahan melewati $[t_{(k)}-\Delta, t_{(k)}]$) = $(r_j - d_j) / r_j$ ... (2.2)

Untuk $t_{(k)} \leq t < t_{(k+1)}$, maka estimator Kaplan-Meier diturunkan berdasarkan peluang bersyarat. Misalkan $t_k \leq t_{k+1}$, maka:

$$\begin{aligned}
F(t_j) &= 1 - S(t_j) \\
&= 1 - P(T > t_j) \\
&= 1 - P(T > t_j, T > t_{j-1}) \\
&= 1 - P(T > t_j \mid T > t_{j-1}) P(T > t_j) \\
F(t_j) &= 1 - P(T > t_j \mid T > t_{j-1}) P(T > t_{j-1} \mid T > t_{j-2}) ... P(T > t_0 = 0)
\end{aligned}$$

Kita asumsikan bahwa dimulainya penelitian semua subjek adalah hidup, jadi $P(\ T > T_0 = 0\ ) = 1$. Maka peluang bersyaratnya adalah:

$$P(T > t_j, T > t_{j-1}) = \frac{r_j - d_j}{r_j} \qquad \text{... (2.3)}$$

Penaksir fungsi *survival* Kaplan-Meier didefinisikan sebagai berikut:

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{r_j - d_j}{r_j} \qquad \text{... (2.4)}$$

Maka estimator fungsi distribusi kumulatif dengan menggunakan metode Kaplan-Meier adalah:

$$\hat{F}(t) = 1 - \prod_{j=1}^{k} \frac{r_j - d_j}{r_j} \qquad \text{... (2.5)}$$

## 2.6 Diagnosa Statistika

### 2.6.1 Pemeriksaan *Outlier*

Dalam sebuah observasi, sering ditemukan unit pengamatan yang *outlier*, dimana selisih antara fungsi distribusi kumulatif *survival* pada data lengkap dengan tidak lengkap (tanpa pengamatan ke-*i*) mempunyai selisih yang tidak sama dengan 0. Untuk melihat apakah pengamatan tersebut *outlier* atau bukan, maka kita dapat lihat dari selisih antara fungsi distribusi kumulatif yang dibentuk dengan data lengkap dan data tidak lengkap (tanpa pengamatan ke-*i*). (Wang *et. al*,. 2014)

Misalkan $\hat{F}_{(i)}(t)$ merupakan fungsi distribusi Kaplan-Meier yang dibentuk setelah penghapusan data ke-*i*, maka:

$$\hat{F}(t) - \hat{F}_{(i)}(t) = \begin{cases} \left[1 - \dfrac{r_i - d_i}{r_i}\right] \displaystyle\prod_{\substack{j=1 \\ (j \neq i)}}^{k} \dfrac{r_j - d_j}{r_j} & t_j \leq t \\ 0, & t_j > t \end{cases}$$

Menurut definisi penaksir Kaplan-Meier, diketahui bahwa:

$$\hat{F}(t) = 1 - \prod_{j=1}^{k} \frac{r_j - d_j}{r_j}$$

$$= 1 - \left\{\frac{r_1 - d_1}{r_1} \times \frac{r_2 - d_2}{r_2} \times \ldots \times \frac{r_{i-1} - d_{i-1}}{r_{i-1}} \times \frac{r_i - d_i}{r_i} \times \frac{r_{i+1} - d_{i+1}}{r_{i+1}} \times \ldots\right\}$$

$$\hat{F}_{(i)}(t) = 1 - \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j}$$

$$= 1 - \left\{\frac{r_1 - d_1}{r_1} \times \frac{r_2 - d_2}{r_2} \times \ldots \times \frac{r_{i-1} - d_{i-1}}{r_{i-1}} \times \frac{r_{i+1} - d_{i+1}}{r_{i+1}} \times \ldots\right\}$$

Jika pengamatan ke-$i$ adalah mati, maka $k$ akan berkurang satu (jumlah interval berkurang satu). Maka diperoleh:

$$\hat{F}(t) - \hat{F}_{(i)}(t) = \left(1 - \prod_{j=1}^{k} \frac{r_j - d_j}{r_j}\right) - \left(1 - \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j}\right)$$

$$= \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j} - \prod_{j=1}^{k} \frac{r_j - d_j}{r_j}$$

$$= \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j} - \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j} \times \left\{\frac{r_i - d_i}{r_i}\right\} \qquad \ldots (2.6)$$

$$= \left[1 - \frac{r_i - d_i}{r_i}\right] \prod_{\substack{j=1 \\ (j \neq i)}}^{k} \frac{r_j - d_j}{r_j}$$

Ketika $t_i > t$, maka $\hat{F} - \hat{F}_{(i)} = 0$

### 2.6.2 Pemeriksaan Data Berpengaruh

Dalam sebuah observasi, biasanya mempunyai pengamatan yang hampir semuanya sama, tidak ada yang menonjol. Ketika kita menghapus unit pengamatan ke-*i*, dan jika fungsi kemungkinannya berbeda antara pengamatan lengkap dengan tidak lengkap (tanpa pengamatan ke-*i*), maka unit pengamatan ke-*i* tersebut berpengaruh.

Wang *et. al* (2014) mengajukan 2 metode yang digunakan untuk mengukur sebuah pengamatan yang berpengaruh terhadap fungsi kemungkinan (*likelihood function*), yaitu:

**1. Perubahan Nilai Kemungkinan (*Likelihood Displacement*)**

Dalam sebuah observasi, ada pengamatan yang berpengaruh terhadap perubahan nilai kemungkinan (*likelihood displacement*). Untuk melihat apakah pengamatan tersebut memang berpengaruh atau tidak, maka kita dapat lihat dari besarnya selisih antara fungsi perubahan nilai kemungkinan yang dibentuk dengan data lengkap dan data tidak lengkap (tanpa pengamatan ke-*i*).

Dalam analisis *survival* ada berbagai kemungkinan data. Pertama adalah kelompok *D* (himpunan individu yang mati atau *death*). Kedua, adalah kelompok *R* (himpunan individu yang tersensor kanan / *right*). Ketiga adalah kelompok *L* (himpunan individu yang tersensor kiri / *left*). Dan keempat adalah kelompok *I* (himpunan individu yang tersensor interval). Maka fungsi kemungkinannya dapat dibentuk sebagai berikut:

$$L \propto \prod_{j\in D} f(x_j) \prod_{j\in R} S(C_r) \prod_{j\in L} (1 - S(C_l) \prod_{j\in I} [S(L_j) - S(R_j)] \qquad \text{... (2.7)}$$

Jika hanya ada individu yang mati dan tersensor kanan, maka fungsi kemungkinannya dapat dibentuk sebagai berikut:

$$L \propto \prod_{j\in D} f(x_j) \prod_{j\in R} S(C_r) \qquad \text{... (2.8)}$$

Untuk yang tersensor kanan, artinya untuk yang $\delta = 0$ maka:

$$P(T, \delta = 0) = P(X > C_r)$$

$$= S(C_r)$$

Untuk δ = 1 maka:

$$P(T, \delta = 0) = f(t)$$

Jika digabungkan menjadi satu persamaan sebagai berikut:

$$P(t, \delta) = [f(t)]^{\delta}[S(t)]^{1-\delta} \quad \text{... (2.9)}$$

Sehingga jika diterapkan pada data *survival*, fungsi kemungkinannya adalah:

$$L = \prod_{j=1}^{n} P(t_j, \delta_j)$$

$$= \prod_{j=1}^{n} \left[f(t_j)\right]^{\delta_j} \left[S(t_j)\right]^{1-\delta_j} \quad \text{... (2.10)}$$

Jadi definisi awal dari fungsi kemungkinan adalah peluangnya, peluang dari masing-masing individu. Karena masing-masing individu tersebut dianggap saling bebas, maka di kalikan.

Diketahui bahwa $f(t_j) = h(t_j) \times S(t_j)$, karena terdapat hubungan dengan persamaan (2.10) maka:

$$L = \prod_{j=1}^{n} \left[h(t_j)\right]^{\delta_j} S(t_j) \quad \text{... (2.11)}$$

atau

$$L(F) = \prod_{j=1}^{k} h(t_j) S(t_j) \prod_{j \in R} S(t_j) \quad \text{... (2.12)}$$

$$L(F_{(i)}) = \prod_{\substack{j=1 \\ (j \neq i)}}^{k} h(t_j) S(t_j) \prod_{j \in R} S(t_j) \quad \text{... (2.13)}$$

Dimaana taksiran kegagalan (hazard) pada interval $t_{(j)} < t < t_{(j+1)}$ adalah

$$\hat{h}(t_j) = \frac{d_j}{r_j(t_{(j+1)} - t_{(j)})}$$

Perubahan nilai kemungkinan adalah metode untuk menghitung pengaruh, seperti yang dikemukakan oleh Cook and Weisberg (1982). Mempertimbangkan pengaruh penghapusan pada data ke-$i$. Maka, perubahan nilai kemungkinan dapat dinyatakan sebagai berikut:

$$LD_i(F) = 2[L(F) - L(F_{(i)})] \quad \text{... (2.14)}$$

Disini akan didefinisikan juga perubahan nilai kemungkinan sebagai berikut:

$$LD_i(F) = -2[ln\,(L(F)) - ln(L(F_{(i)}))] \quad \text{... (2.15)}$$

Perubahan nilai kemungkinan tersebut berdistribusi Chi-kuadrat dengan derajat bebas 1. Maka nilai batas dengan α = 0,05 bagi pemeriksaan pengamatan berpengaruh adalah sebesar $\chi^2_{(1\,;0{,}05)}$ = 3,84146.

**2. Statistik Rasio Kemungkinan (*Likelihood Ratio Statistic*)**

Ada alternatif lain di dalam melihat pengamatan yang berpengaruh yaitu dengan mengunakan statistik rasio kemungkinan (*likelihood ratio statistic*) dimana pada bagian ini akan dibandingkan antara fungsi rasio kemungkinan pada data lengkap dan tidak lengkap (tanpa pengamatan ke-*i*). Begitu pula kita mendefinisikan statistik rasio kemungkinan untuk data tersensor kanan berdasarkan *estimator* Kaplan-Meier sebagai berikut:

$$R_i = \frac{L(F_{(i)})}{L(F)} = \frac{\prod_{\substack{j=1 \\ (j\neq i)}}^{k} h(t_j)S(t_j)\ \prod_{j\in R} S(t_j)}{\prod_{j=1}^{k} h(t_j)S(t_j)\ \prod_{j\in R} S(t_j)} \quad \text{... (2.15)}$$