II. TINJAUAN PUSTAKA

2.1 Analisis Survival

Analisis survival (*survival analysis*) atau analisis kelangsungan hidup bertujuan menduga probabilitas kelangsungan hidup, kekambuhan, kematian, dan peristiwa-peristiwa lainnya sampai pada periode waktu tertentu. Dalam menentukan waktu survival, terdapat tiga elemen yang harus diperhatikan yaitu

a. *Time origin or starting point* (titik awal) adalah waktu dimulainya suatu penelitian.
b. *Ending event of interest* (kejadian akhir) adalah kejadian yang menjadi inti dari penelitian
c. *Measurement scale for the passage of time* (skala ukuran untuk berlalunya waktu).

Perbedaan antara analisis survival dengan analisis statistik lainnya adalah adanya data tersensor. data dikatakan tersensor jika pengamatan waktu survival hanya sebagian, tidak sampai waktu kejadian.

Menurut Lee (1992), jika Y melambangkan waktu survival dan mempunyai fungsi kepekatan peluang $f(y)$, maka fungsi sebaran kumulatif dinyatakan sebagai

$$F(y) = P(Y \leq y) = \int_0^y f(y)dt$$

yang merupakan peluang kejadian gagal sebelum waktu y.

Dalam bidang kesehatan data ini diperoleh dari suatu pengamatan terhadap sekelompok atau beberapa kelompok individu dan dalam hal ini adalah pasien, yang diamati dan dicatat waktu terjadinya kegagalan dari setiap individu. Kegagalan yang dimaksudkan antara lain adalah kematian karena penyakit tertentu, keadaan sakit yang terulang kembali setelah pengobatan atau munculnya penyakit baru. Apabila kegagalan yang diamati adalah terjadinya kematian pada pasien maka waktu survival yang dicatat antara lain sebagai berikut :

a. Selisih waktu mulai dilakukannya pengamatan sampai terjadinya kematian dan data tersebut termasuk data tidak terpotong (*uncensored data*).
b. Jika waktu kematiannya tidak diketahui, maka memakai selisih waktu mulai dilakukannya pengamatan sampai waktu terakhir penelitian dan data tersebut termasuk data terpotong (*censored data*).

2.2 Fungsi Survival

Menurut Lee (1992), jika u variabel random yang menotasikan waktu bertahan dari seorang individu, maka $S(t)$ adalah probabilitas bahwa T lebih besar dari t. Dalam statistik fungsi distribusi kumulatif $F(t)$ didefinisikan :

$$F(t) = P_r\ (T \leq t)$$

$$= \int_{-\infty}^{t} f(u)\ du$$

Karena $t > 0$ maka $F(t) = \int_0^t f(u) du$

$S(t) = P_r$ (suatu individu bertahan lebih dari t)

$$= P_r(T > t)$$

$$= \int_{t}^{\infty} f(u)du$$

Dari definisi fungsi distribusi kumulatif $F(t)$ dari T, maka

$S(t) = 1 - P\ (suatu\ individu\ gagal\ sebelum\ waktu\ t)$

$S(t) = 1 - F(t)$

2.3 Data Tersensor

Suatu data dikatakan tersensor jika lamanya hidup seseorang yang ingin diketahui atau diobservasi hanya terjadi pada periode waktu yang telah ditentukan (interval pengamatan), sedang info yang ingin diketahui tidak terjadi pada interval tersebut. Dengan demikian kita tidak memperoleh informasi apapun yang diinginkan selama interval pengamatan.

Menurut Collect (1997), data survival tidak memenuhi syarat prosedur standar statistika yang digunakan pada analisis data. Alasan pertama karena data survival biasanya berdistribusi tidak simetris. Model histogram waktu survival pada sekelompok individu yang sangat sama akan cenderung "positive skewed", oleh karena itu histogram akan semakin miring ke kanan sesuai dengan interval waktu dengan jumlah pengamatan terbesar, sehingga tidak ada alasan untuk mengasumsikan bahwa data survival berdistribusi normal.

Ada empat jenis penyensoran yaitu sensor kanan *(right censoring),* sensor kiri *(left censoring)* , sensor selang *(interval censoring),* dan sensor acak *(random censoring). Right cencoring,* terjadi jika individu yang diamati masih tetap hidup pada saat waktu yang telah ditentukan. *Left cencoring,* terjadi jika semua informasi yang ingin diketahui dari seorang individu telah dapat diperoleh pada awal studi. *Interval cencoring,* jika informasi yang dibutuhkan telah dapat diketahui pada kejadian peristiwa didalam selang pengamatan. *Random cencoring* terjadi apabila individu yang di amati meninggal karena sebab lain, bukan disebabkan dari tujuan utama penelitian (Klein dan Moeschberger, 1997).

Dalam contoh yang di ilustrasikan oleh Klein dan Moeschberger (1997), misalkan dalam sebuah penelitian untuk menentukan sebaran penggunaan ganja dikalangan anak laki-laki di sebuah SMA di California. Dengan mengajukan pertanyaan “kapan anda menggunakan ganja untuk pertama kalinya”. Ternyata ada beberapa anak menjawab “saya tidak pernah menggunakan ganja”, dengan demikian anak tersebut mengalami kejadian tersensor kanan. Dan jika ada anak yang menjawab “saya pernah menggunakanya, tetapi saya tidak tahu tepatnya kapan pertama kali menggunakanya”, pada kasus ini anak tersebut mengalami kejadian tersensor kiri dikarenakan waktu awal dia menggunakan ganja tidak diketahui. Pada kasus lain, misalkan untuk mengestimasi distribusi dari beberapa tikus yang diberikan zat karsinogen pada makanannya, dilakukan studi selama 10 bulan kepada 10 tikus dan penelitian dilakukan setiap akhir tahun, jika 2 dari 8 tikus tewas karena kanker pada bulan ke-5 dan ke-7, maka 2 tikus tersebut mengalami kejadian sensor selang, karna tidak didapat informasi kapan tepatnya tikus tersebut tewas karena kanker. Dan jika ada 1 dari 10 tikus tersebut meninggal karena terinjak (tewas bukan karena penilitian utam) bukan karena terkena kanker, maka tikus tersebut mengalami sensor acak.

2.4 Tipe-Tipe Penyensoran

Jenis penyensoran dapat dibagi lagi menjadi tipe-tipe penyensoran. Menurut Johnson (1982), tipe-tipe penyensoran terdiri dari :

1. Penyensoran Tipe I

Pada penyensoran sebelah kanan tipe I, penelitian diakhiri apabila waktu pengamatan yang ditentukan tercapai. Jika waktu pengamatan sama untuk semua unit maka dikatakan penyensoran tunggal. Jika waktu pengamatan untuk setiap unit berbeda maka dikatakan penyensoran ganda.

Pada penyensoran sebelah kiri tipe I, pengamatan dilakukan jika telah melampaui awal waktu yang ditentukan. Karakteristik penyensoran tipe I adalah bahwa kegagalan adalah acak.

2. Penyensoran Tipe II

Pada penyensoran tipe II, pengamatan diakhiri setelah sejumlah kegagalan yang telah ditetapkan diperoleh, atau dapat dikatakan banyaknya kegagalan adalah tetap dan waktu pengamatan adalah acak.

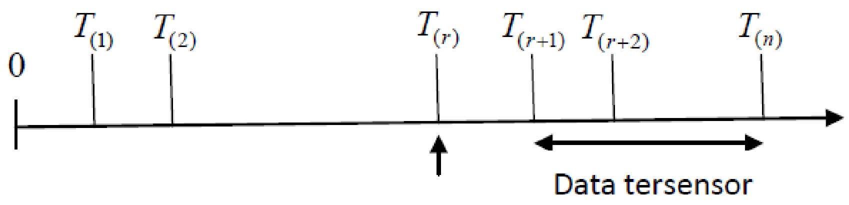
Pada sensor kanan jenis II, jumlah individu pada saat awal ditentukan dan waktu penelitian ditentukan sampai terjadinya kematian dengan jumlah tertentu.

3. Penyensoran Maju (*Progressive Censoring*)

Pada penyensoran maju, suatu jumlah yang ditentukan dari unit-unit bertahan dikeluarkan dari penelitian berdasarkan kejadian dari tiap kegagalan terurut. Secara konseptual, hal ini sama dengan suatu praktek yang dikenal sebagai *sudden-death testing*, dimana tes secara serempak memuat beberapa pengetesan dan apabila terjadi kegagalan pertama maka seluruh pengetesan dianggap gagal.

2.5 Model *Survival* Data Tersensor KananTipe II

Dalam data tersensor kanan tipe 2, terdapat r pengamatan dari n sampel yang diamati, dan eksperimen akan dihentikan setelah kegagalan $ke-r$ yang terjadi sebelum waktu t_i. Data terdiri dari r tahan hidup terkecil $T_{(1)} \leq T_{(2)} \leq T_{(3)} \leq \cdots \leq T_{(r)}$ dari sampel random yang terdiri dari n tahan hidup $T_1, T_2, T_3, \dots, T_n$, seperti iliustrasi pada Gambar 2.1.



Gambar 2.1. Ilustrasi Model Tersensor Kanan

Misalkan T merupakan variabel random dari n individu yang diamati, $f(t_1)$ merupakan fungsi densitas peluang dari variabel random individu $ke-1$, $f(t_2)$ merupakan fungsi densitas peluang dari variabel random individu $ke-2$, dan seterusnya hingga $f(t_r)$ untuk variabel random individu $ke-r$. Individu yang gagal, yaitu individu $ke-1$ sampai individu $ke-r$ masing-masing sebanyak satu komponen. Sedangkan individu yang masih bertahan melebihi kegagalan dari individu $ke-r$ dituliskan sebagai $T_{r+1}, T_{r+2}, T_{r+3}, \dots, T_n$ sebanyak $n-r$. Sampel random berukuran n dengan kegagalan r ini mengikuti distribusi multinomial, sehingga terdapat $\frac{n!}{1!1!\dots1!(n-r)!}$ urutan yang mungkin terjadi dari n pengamatan.

Fungsi densitas peluang bersama dari $T_1, T_2, T_3, \dots, T_n$dari data yang diamati, dapat ditulis sebagai berikut:

$$f(t_1, t_2, \dots, t_n) = \frac{n!}{(n-r)!} f(t_1) f(t_2) \dots f(t_r) [P(T_{r+1} \geq t_r) \dots P(T_n \geq t_r)]$$

$$= \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f(t_i) \right] \left[\left(1 - P(T_{r+1} < t_r)\right) \dots \left(1 - P(T_n < t_r)\right) \right]$$

$$= \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f(t_i) \right] \left[\left(1 - F(t_r)\right) \dots \left(1 - F(t_r)\right) \right]$$

$$= \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f(t_i) \right] [1 - F(t_r)]^{n-r}$$

$$= \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f(t_i) \right] [S(t_r)]^{n-r}$$

Fungsi densitas peluang bersama data tersensor kanan dari $t_1, t_2, \dots, t,$ untuk $r < n$ dengan order statistik adalah:

$$f(t_1, t_2, \dots, t_r) = \frac{n!}{(n-r)!} [\prod_{i=1}^{r} f(t_i)][S(t_r)]^{n-r} \quad (2.1)$$

(Klein dan Moeschberger, 1997)

2.6 Distribusi Weibull

Distribusi Weibull dikembangkan oleh W. Weibull pada awal tahun 1950. Distribusi Weibull adalah salah satu distribusi yang penting pada teori *reabilty*. Distribusi Weibull sangat luas digunakan untuk analisa kehilangan performasi pada sistem kompleks di dalam sistem engineering. Secara umum, distribusi ini dapat digunakan untuk menjelaskan data saat waktu menunggu hingga terjadi kejadian dan untuk menyatakan berbagai fenomena fisika yang berbeda-beda. Dengan demikian, distribusi ini dapat diterapkan pada analisa resiko karena dapat menduga umur pakai (*life time*) komponen. Distribusi Weibull merupakan salah satu jenis distribusi kontinu yang sering digunakan, khususnya dalam bidang keandalan dan statistik karena kemampuannya untuk mendekati berbagai jenis sebaran data.

Distribusi Weibull sering digunakan untuk memodelkan waktu kegagalan dari banyak sistem fisik. Parameter dalam distribusi memungkinkan fleksibilitas untuk memodelkan sistem dengan jumlah kegagalan bertambah terhadap waktu, berkuran terhadap waktu atau tetap konstan terhadap waktu.

Menurut Al-Kanani (2011), fungsi kepekatan peluang dari suatu peubah acak Weibull (μ, β) dengan parameter $\mu \geq 0$ dan $\beta \geq 0$, adalah sebagai berikut:

$$f(x; \mu, \beta) = \left(\frac{\beta}{\mu}\right)\left(\frac{x}{\mu}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\mu}\right)^{\beta}\right] \quad ; x > 0,$$

Distribusi Weibull dengan parameter μ dan β dapat dinotasikan sebagai berikut $\boldsymbol{WE}(\mu, \beta)$, dengan $\mu = parameter\ skala$ dan $\beta = parameter\ bentuk.$

Dengan nilai Mean dan Varian berturut-turut sebagai berikut:

$$E(x) = \mu\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]$$

$$V(x) = \mu^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)^2\right)\right]$$

Dan *commulatife density functions* (*cdf*) dari Distribusi Weibull, yaitu :

$$\mathrm{F}(x;\mu,\beta) = 1 - \exp\left[-\left(\frac{x}{\mu}\right)^{\beta}\right]$$

Dan fungsi *survivalnya* adalah

$$S(t;\mu,\beta) = \exp\left[-\left(\frac{t}{\mu}\right)^{\beta}\right]$$

2.7 Metode Kemungkinan Maksimum (*maksimum likelihood estimation method*)

Metode kemungkinan maksimum adalah metode untuk menduga satu sebaran dengan memilih dugaan-dugaan yang nilai-nilai parameternya diduga dengan memaksimalkan fungsi kemungkinanya, metode kemungkinan maksimum merupakan salah satu metode yang paling sering digunakan untuk mencari nilai estimasi dari suatu parameter.

Menurut Nar Herrhyanto (2003), misalkan X adalah peubah acak kontinu atau diskrit dengan fungsi kepekatan peluang $f(x;\theta)$, dengan θ adalah salah satu sampel yang tidak diketahui. Misalkan $x_1, x_2, \dots, x_n$ merupakan sampel acak berukuran n maka fungsi kemungkinan (likelihood function) dari sampel acak itu adalah:

$$L(\theta) = f(x_1;\theta)\, f(x_2;\theta) \dots f(x_n;\theta)$$

Dalam hal ini, fungsi kemungkinan adalah fungsi dari parameter yang tidak diketahui θ. Biasanya untuk mempermudah penganalisaan, fungsi kemungkinan $L(\theta)$ diberi log natural (ln). Penduga kemungkinan maksimum dari θ adalah nilai θ yang memaksimalkan fungsi $L(\theta)$.

2.8 Metode Newton-Raphson

Kebanyakan persoalan model matematika dalam bentuk yang rumit tidak dapat diselesaikan dengan metode analitik yang sudah umum untuk mendapatkan solusi eksak. Bila metode analitik tidak dapat lagi diterapkan, maka solusi dari persoalan model matematika tersebut masih dapat diselesaikan dengan menggunakan metode numerik.

Menurut Atkinson (1993) dalam metode numerik, pencarian akar $f(x) = 0$ dilakukan dengan iterasi. Diantara semua metode akar, metode Newton-Raphsonlah yang paling terkenal dan paling banayak dipakai dalam terapan sains dan rekayasa. Metode ini paling disukai karena tingkat konvergensinya paling cepat diantara metode lainnya

$$x_{r+1} = x_r - \frac{f(x_r)}{f'(x_r)},\ f'(x_r) \neq 0$$

Dimana

x_r adalah nilai awal

$f(x_r)$ adalah fungsi yang ingin diketahui nilainya

$f'(x_r)$ adalah turunan pertama dari $f(x_r)$

2.9 Program *R*

R adalah perangkat lunak bebas untuk komputasi statistik dan grafik. Merupakan proyek GNU *General Public License Free Software Foundation* yang mirip dengan bahasa S yang dikembangkan di Bell Laboratories oleh Jhon Chambers dan rekan. R menyediakan berbagai statistik seperti linear dan nonlinear *modeling*, pengujian analisis klasik, analisis *time-series*, klasifikasi dan lainnya. Sebuah rangkaian perangkat lunak yang digunakan untuk manipulasi data, perhitungan, dan tampilan grafik yang mencakup antara lain sebagai berikut :

a. Penanganan data yang efektif dan penyimpanan data.
b. Rangkaian operator untuk perhitungan array dalam matriks tertentu.
c. Fasilitas grafik untuk analisis data dan menampilkan baik pada layar maupun *hardcopy*.
d. Bahasa pemprograman yang sederhana, berkembang dengan baik dan efektif. (r-project.org)