BAB II

KONSEP DASAR

2.1 Konsep Analisis Ketahanan Hidup

Survival analysis (ketahanan hidup) berkaitan erat dengan survival time (waktu ketahanan hidup). Sebab dengan diketahuinya waktu ketahanan hidup suatu individu bisa diketahui berapa besar kemungkinan ketahanan hidup individu tersebut oleh suatu perlakuan.

Untuk mengetahui ketahanan hidup dapat diperoleh dengan mengestimasi dari distribusi fungsi-fungsi ketahanan hidup.

Distribusi waktu ketahanan hidup biasanya dapat dinyatakan oleh tiga fungsi yang ekuivalen yaitu :

1. Fungsi ketahanan hidup
2. Fungsi probabilitas densitas
3. Fungsi Hazard

Waktu ketahanan hidup (T) merupakan variabel random non negatif yang mewakili ketahanan hidup dari individu-individu dalam populasi yang merupakan variabel random kontinu dalam interval $[0,\infty)$ atau ketahanan hidup pada waktu t dengan $t > 0$.

2.1.1 Fungsi Ketahanan Hidup

Fungsi survival adalah probabilitas bahwa suatu individu akan tetap hidup sampai waktu t $(t > 0)$. Jadi jika u variabel random yang menotasikan waktu bertahan hidup dari seorang individu, maka S(t) adalah probabilitas bahwa T lebih besar dari t. Dalam Statistik fungsi kumulatif distribusi F (t) didefinisikan :

$$F(t) = Pr(T \le t) = \int_{-\infty}^{t} f(u)\, du$$

Karena t > 0 maka $F(t) = \int_{0}^{t} f(u)\, du$ ....(2.1)

Jadi

$$S(t) = Pr(\text{individu tahan hidup lebih dari } t) = Pr(T > t) = \int_{t}^{\infty} f(u)\, du$$

Dari definisi fungsi distribusi kumulatif F(t) dapat diperoleh hubungan

S(t) = 1 - P (Individu gagal sebelum waktu t)

$S(t) = 1 - F(t)$ .....(2.2)

Fungsi ketahanan hidup adalah fungsi tidak naik yang bersifat :

$S(t) = 1$ untuk $t = 0$

$S(t) = 0$ untuk $t = \infty$

Yaitu bahwa probabilitas suatu individu bertahan hidup pada waktu 0 adalah 1 dan probabilitas bertahan hidup sampai waktu mendekati tak berhingga adalah nol.

2.1.2 Fungsi Probabilitas Densitas

Jika f (t) adalah fungsi kepadatan probabilitas (fkp) yang didefinisikan sebagai limit probabilitas individu gagal dalam interval periodik t sampai $t + \Delta t$ perunit lebar Δt atau lebih mudahnya ialah probabilitas kegagalan dalam sebuah interval pendek perunit waktu yang didefinisikan sebagai berikut :

$$f(t) = \lim_{\Delta t \to 0} \frac{Pr\ (\text{individu mati dalam interval } (t, t+\Delta t)}{\Delta t} \quad ..(2.3)$$

Fungsi densitas kematian bersifat

1. f(t) ialah fungsi non negatif

   $f(t) \geq 0$, untuk semua $t \geq 0$

   $f(t) = 0$ untuk semua $t < 0$

2. Luas daerah antara kurva densitas dengan sumbu t sama dengan 1 yaitu : $\int_0^\infty f(t)\, dt = 1$

2.1.3 Fungsi Hazard

Fungsi hazard dari waktu ketahanan hidup T merupakan Probabilitas bersyarat bahwa individu akan mati pada interval $[t, t + \Delta t]$ jika diketahui individu tetap hidup sampai saat t adalah

$$\frac{F(t + \Delta t) - F(t)}{S(t)}$$

Dengan membagi kuantitas ini dengan Δt dan mengambil limit untuk $\Delta t \longrightarrow 0$ maka akan diperoleh rumus yang dikenal dengan Hazard Rate atau $\lambda(t)$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t \cdot S(t)}$$

$$= \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.4)$$

Karena $f(t) = F'(t)$ maka menjadi

$$\lambda(t) = \frac{F'(t)}{S(t)} = \frac{1}{1 - F(t)} \frac{dF(t)}{dt}$$

$$\int_0^t \lambda(u)\, du = \int_0^t \frac{dF(u)}{1 - F(u)} \frac{du}{du}$$

$$= -\log [1 - F(u)] \Big|_c^t$$

$$\exp [-\int_0^t \lambda(u)\, du = [1 - F(t)]$$

Dari (2.2) dan (2.4)

$$S(t) = \exp [-\int_0^t \lambda(u)\, du]$$

$$f(t) = \lambda(t)\ .\ S(t)$$

$$f(t) = \lambda(t)\ .\ \exp [-\int_0^t \lambda(u)\, du] \qquad \text{.................(2.5)}$$

Dengan demikian jika Hazard Rate dari suatu distribusi tahan hidup diketahui maka f(t), F(t) dan S(t) dapat dicari.

2.2 Konsep Metode Nonparametrik

Metode estimasi dan uji hipotesis dikatakan nonparametrik jika metode ini berlaku tanpa asumsi tertentu tentang bentuk distribusi populasinya.
Metode nonparametrik biasanya memanfaatkan beberapa aspek sederhana dari data sampel seperti tanda aljabar observasi, hubungan urutan atau frekwensi katagori.
Distribusi nol suatu statistik nonparametrik dapat ditentukan tanpa mengingat bentuk distribusi populasinya karena alasan ini maka prosedur inferensi nonparametrik juga disebut distribusi bebas (free distribusi).

Uji nonparametrik adalah suatu uji yang tidak membutuhkan model parameter khusus dari populasi yang diamati. Beberapa asumsi yang berhubungan erat dengan uji statistik nonparametrik adalah bahwa pengamatan tersebut bebas dan variabel yang diamati kontinu, tetapi asumsi yang dibuat adalah lebih lemah bila dibandingkan dengan uji parametrik. Oleh karena itu uji nonparametrik tidak membutuhkan suatu pengukuran dengan tingkat ketelitian yang tinggi seperti uji parametrik.
Biasanya uji nonparametrik dipakai untuk menganalisa data dalam skala ordinal, nominal serta interval/rasio.

Apabila akan melakukan analisa data ketahanan hidup dari sekumpulan data, baik data yang diperoleh sebagai hasil eksperimen maupun data sekunder, maka yang pertama harus diketahui model distribusi dari sekumpulan data tersebut. Bila model distribusinya berkaitan erat dengan model distribusi tahan hidup pada umumnya bisa dicari dengan metode parametrik. Tetapi bila belum diketahui model distribusinya dengan pasti maka digunakan dengan metode nonparametrik.

Untuk mengetahui model distribusi data dengan nonparametrik adalah dengan mencari Hazard Rate, fungsi probabilitas (Failure rate) dan fungsi survival (Reliabilitas) untuk 1 sampel sedangkan untuk dua sampel sampai k sampel dapat dilakukan dengan pengujian hipotesis.

2.2.1 Pengujian Hipotesa

Definisi 2.1

Suatu hipotesa statistik adalah suatu pengamatan dari distribusi dengan satu atau lebih random variabel. Jika hipotesa statistik dengan distribusi dapat diuraikan secara lengkap disebut hipotesa statistik sederhana, dan jika tidak dapat disebut hipotesa majemuk.

Definisi 2.2

Suatu uji dari suatu hipotesa statistik adalah suatu pernyataan saat nilai sampel percobaan yang berguna untuk menerima atau menolak hipotesa yang digunakan atau diinginkan.

Definisi 2.3

H_o disebut hipotesa nol yang merupakan hipotesa yang akan diuji.

Definisi 2.4

H_1 disebut hipotesa alternatif yang merupakan tandingan dari H_o maksudnya jika H_o ditolak maka H_1 akan diterima.

Definisi 2.5

Kesalahan jenis I (α) adalah penolakan hipotesa nol (H_o) yang benar. Biasanya α yang digunakan 0,01 atau 0,05.

Definisi 2.6

Kesalahan jenis II (β) adalah penerimaan hipotesa nol (Ho) yang salah.

Sesudah memilih hipotesa tertentu yang tampaknya penting dalam suatu percobaan, kita mengumpulkan data emperis yang harus menghasilkan informasi langsung mengenai dapatnya hipotesa ditolak atau tidak ditolak. Keputusan kita mengenai arti data itu mungkin membuat kita mempertahankan, merevisi atau menolak hipotesa tersebut. Obyektivitas harus ditekankan sebab salah satu yang dituntut dari metode ilmiah adalah bahwa seseorang harus sampai pada kesimpulan ilmiah melalui metode-metode yang diketahui umum dan dapat diulangi oleh peneliti lain yang berkompeten.

Prosedur yang biasanya diikuti terdiri dari beberapa langkah :

1. Menyatakan hipotesa nol (Ho)
2. Memilih salah satu test statistik (dengan model statistik yang berkaitan) untuk menguji Ho. Dari beberapa test yang dipergunakan pada suatu rancangan penelitian tertentu memilih test yang modelnya paling mendekati persyaratan penelitian (sehubungan dengan asumsi-asumsi yang memperbolehkan penggunaan test itu) dan syarat pengukurannya dapat dipenuhi oleh ukuran-ukuran yang dipergunakan dalam penelitian.

3. Menetapkan suatu tingkat signifikansi (α) dari besarnya sampel N.
4. Berdasarkan 1,2 dan 3 diatas menentukan daerah penolakan (region of rejection)
5. Menghitung harga statistik itu dengan menggunakan data yang diperoleh dari sampel-sampelnya. Jika harga itu masuk dalam daerah penolakan, keputusannya adalah menolak Ho, jika harga itu diluar deerah penolakan keputusannya adalah Ho tidak dapat ditolak pada tingkat signifikansi yang sudah dipilih.

2.2.2 Fungsi Distribusi Kumulatif Sampel

Maksud dan tujuan cara sampling dari beberapa distribusi adalah untuk membuat inferensi tentang distribusi sampel atau populasi. Kita dapat mengestimasi fungsi distribusi kumulatif yang tidak diketahui dengan menggunakan sampel atau mengestimasi fungsi distribusi kumulatif yang merupakan statistik berurut.

Definisi 2.7 (Statistik Berurut)

Misalkan $t_1, t_2, \ldots, t_n$ menunjukkan suatu sampel random ukuran n dari suatu fungsi distribusi kumulatif F(t). Maka $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(n)}$ dimana T_i adalah t_i yang disusun dalam urutan besaran yang bertambah dan didefinisikan sebagai statistik berurut yang bersesuaian dengan sampel random $t_1, t_2, \ldots, t_n$.

Sedangkan fungsi distribusi kumulatif sampel didefinisikan oleh definisi dibawah ini.

Definisi 2.8 (Fs distribusi kumulatif sampel)

Misalkan t_1, t_2, ... ,t_n menunjukkan suatu sampel random dan suatu fungsi distribusi kumulatif F(t) dan ambil T_1, T_2, ... , T_n menunjukkan hubungan statistik berurut.Fungsi distribusi kumulatif sampel ditunjukkan oleh :

$$F_n(t) = \frac{1}{n} \text{ x banyaknya } \quad T_i \le t$$

2.2.3 Metode Rank Statistik

Rank statistik merupakan salah satu teknik penyelesaian yang biasa dipakai dalam metode statistik nonparametrik.

Definisi 2.9

Rank statistik suatu random variabel adalah suatu himpunan yang anggota-anggotanya menunjukkan posisi relatif dari suatu pengamatan (observasi).

Bila X_1, X_2, ..., X_n sampel berukuran n dan R_i notasi untuk rank statistik ke-i maka :

R_i = 1 bila X_i harga terkecil diantara X_1, X_2, ... , X_n.

R_i = 2 bila X_i harga terkecil kedua diantara X_1, X_2,...,X_n.

.

.

R_i = n bila X_i harga terkecil ke n dari X_1, X_2, ... , X_n.

$R_i \le R_j$ bila $X_i \le X_j$

Jadi rank statistik diperoleh bila observasi diganti dengan nilai urutan posisi relatifnya.

Bila pada suatu susunan data terdapat 2 atau lebih observasi dengan nilai yang sama ("ties") maka ranknya adalah rata-rata rank bila dipandang tanpa ties.

Contoh :

Data terdiri atas 10,7,-2,-1,7,3

Tabel 1.1 Cara Menentukan Rank

	X_i	R_i dipandang tanpa ties	R_i yang sebenarnya
1	10	6	6
2	7	4	4,5
2	-2	1	1
3	-1	2	2
4	7	5	4,5
5	3	3	3

$$\left.\begin{array}{l} X_2 = 7 \\ X_5 = 7 \end{array}\right] \text{ ties}$$

Rank X_2 bila dipandang tanpa ties $R_2 = 4$ atau $R_5 = 5$

Rank X_5 bila dipandang tanpa ties $R_5 = 5$ atau $R_2 = 4$

$$R_2 = R_5 = \frac{4 + 5}{2} = 4,5$$

Metode rank statistik tidak hanya berguna untuk menganalisa data yang berbentuk "bilangan" tetapi dapat pula untuk menganalisa data yang berbentuk pernyataan. Misal : baik lebih baik, terbaik dsb , asal pada data tersebut dapat ditentukan ranknya.

Karena dalam skripsi ini membicarakan ketahanan hidup suatu individu maka bila di dapat waktu kegagalan yang sama (tie) maka rank diganti dengan rank terbesar dari waktu ketahanan hidup yang sama tersebut.

2.3 Beberapa Type Penyensoran

Untuk mendapatkan data uji ketahanan hidup biasanya orang melakukan eksperimen. Dalam eksperimen tersebut ada beberapa metode yang dapat dilakukan sehingga macam data yang diperoleh juga berbeda dari satu metode dengan metode yang lain.

Ada berapa eksperimen yaitu :

1. Eksperimen yang menghasilkan data sukses/gagal
2. Eksperimen berdasarkan waktu (survival time) yaitu menghitung hasil dari suatu operasi (eksperimen) berdasarkan waktu sebenarnya.

Dalam eksperimen biasanya akan timbul sensor. Dan sensor ini ada bermacam-macam:sensor type I, sensor type II, sensor random, sensor kanan, sensor kiri, sensor progresif dan lain-lain. Sebuah pengamatan sensor hanya memuat sebagian informasi tentang variabel yang dikehendaki.

2.3.1 Sampel Lengkap

Dalam sampel lengkap eksperimen akan dihentikan jika semua komponen yang diuji telah mati gagal. Cara ini mempunyai satu keuntungan yaitu dapat diperoleh observasi berurut dari suatu komponen yang akan diuji.

2.3.2 Sensor Type I (Sensor Waktu)

Dalam sensor type I, eksperimen akan dihentikan jika telah dicapai waktu tertentu (waktu sensor t_c) Andai T_1, T_2, ... , T_n adalah sampel random dari distribusi ketahanan hidup, mengamati Y_1, Y_2, ... , Y_n dimana:

$$Y_i = \begin{cases} T_i \text{ jika } T_i \le t_c \\ t_c \text{ jika } T_i > t_c \end{cases}$$

2.3.3 Sensor Type II

Suatu sampel dikatakan tersensor type II bila eksperimen dihentikan setelah kegagalan ke r telah diperoleh. Jika T_1, T_2, ... T_r adalah observasi berurut dari sampel berukuran n yaitu $T_{(1)} < T_{(2)} < \ldots\ldots < T_{(n)}$ jadi urutan pengambilan sampel seluruhnya adalah :

$$Y_{(1)} = T_{(1)}$$

.

$$Y_{(r)} = T_{(r)}$$

$$Y_{(r+1)} = T_{(r)}$$

$$Y_{(n)} = T_{(r)}$$

Sehingga terlihat bahwa untuk $r < n$ diperoleh pengamatan individu Y sampai ke n adalah sama dengan pengamatan individu T sampai ke r.

2.3.4 Sensor Random

Penyensoran random adalah penyensoran yang terjadi pada penelitian yang dilakukan dalam waktu berlainan dimana data ketahanan hidup yang diperoleh merupakan data tersensor. Sehingga informasi tentang sensor untuk setiap data selalu dinyatakan secara eksplisit.

Misalkan C_i adalah waktu sensor yang berhubungan dengan $T_{(i)}$ data yang diperoleh adalah (Y_i , δ_i)

Dimana :

$$Y_i = \min (T_i , C_i)$$

$$\delta_i = I(T_i \le C_i) = \begin{cases} 1, \text{jika } T_i \le C_i : T_i \text{ tidak tersensor} \\ 0, \text{jika } T_i > C_i : T_i \text{ tersensor} \end{cases}$$

Tipe sensor random ini sering timbul pada aplikasi dibidang medis. Pada suatu percobaan medis sering terjadi pasien yang dipelajari memasuki studi dalam waktu yang berlainan ,dan setiap pasien diberikan suatu terapi .
Dari percobaan ini akan diperoleh waktu hidup setiap pasien tetapi penyensoran mungkin saja terjadi karena disebabkan berbagai hal ,antara lain :

1. Loss to follow up.

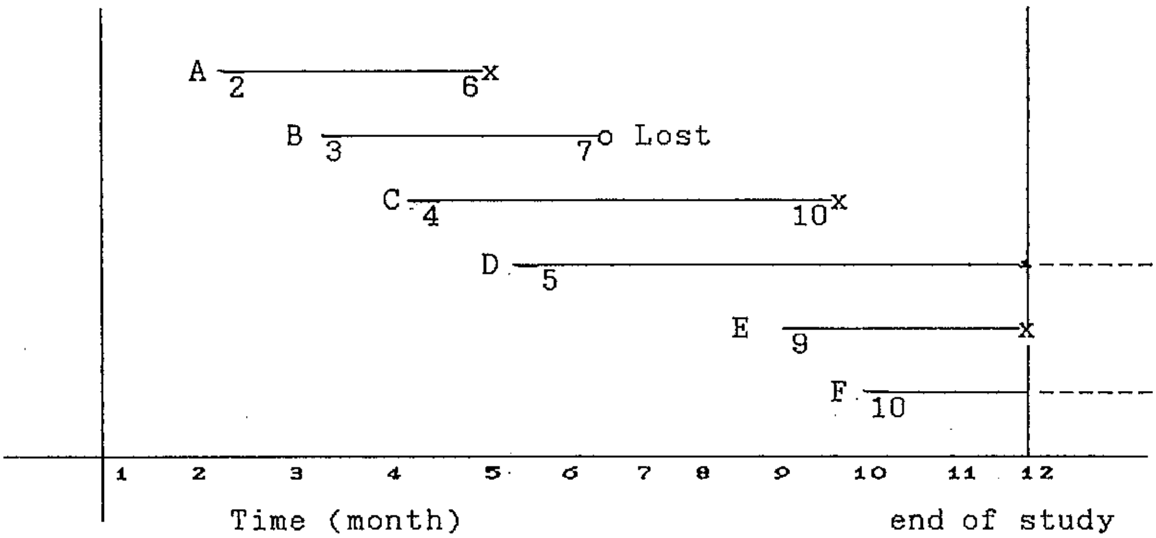
   Pasien berpindah tempat tinggal sehingga tak bisa ditemui oleh peneliti
2. Drop out.

   Pasien tetap ditempat tetapi menolak untuk melanjutkan terapi lagi
3. Terbatasnya waktu penelitian.

Sebagai ilustrasi data tersensor secara random diberikan contoh berikut :
Diadakan penelitian medis pada pasien penyakit leukimia untuk mengetahui keampuhan suatu pengobatan baru .

Penelitian ini melibatkan enam orang pasien yang memasuki masa penelitian pada waktu yang berbeda, dimana masa penelitian ini adalah satu tahun.

Setelah masa penelitian habis terlihat hasilnya seperti gambar.



Gambar : 1.1 Penelitian medis pasien leukimia yang terkena sensor

Disini terlihat pasien A, C dan E tidak tersensor dan pasien B, D dan F tersensor. Disini pasien B tersensor karena ia meninggalkan penelitian sebelum masa penelitian habis. Waktu ketahanan hidup untuk masing-masing pasien dapat dihitung sebagai berikut :

Untuk pasien A. Tidak tersensor

$T = ta^{\bullet} - ta$

$= 6 - 2 = 4$

$C = t^{*} - ta$

$= 12 - 2 = 10$

Disini $T < C$, sehingga data yang diperoleh adalah (4,1) jadi pasien A tidak tersensor.

Untuk pasien B, tersensor

$$T = tb^{\bullet} - tb$$

$$= 7+ - 3 = 4+$$

$$C = t^{*} - tb$$

$$= 12 - 3 = 9$$

Disini $T < C$, sehingga data yang diperoleh adalah (4+, 0) jadi pasien B tersensor

Untuk pasien D, tersensor

$$T = td^{\bullet} - td$$

$$= 12+ - 5 = 7+$$

$$C = t^{*} - td$$

$$= 12 - 5 = 7$$

Disini $T > C$ sehingga data yang diperoleh (7+,0), jadi pasien D tersensor.

Data ketahanan hidup untuk pasien C, E dan F juga dicari dengan cara yang sama sehingga diperoleh waktu ketahanan hidup untuk seluruh pasien berturut-turut 4, 4+, 6, 7+, 3 dan 2+

2.4 Estimasi Produk Limit

Dalam analisis ketahanan hidup banyak dijumpai data tersensor maka diperlukan metode penaksiran khusus. Taksiran produk limit adalah salah satu metode penaksiran nonparametrik untuk fungsi ketahanan (survival fungsion) yang dapat dipakai untuk data ketahanan dimana data ketahanan terdiri dari observasi yang tersensor. Metode ini pertama kali dikembangkan oleh Kaplan dan Meier (1958). Metode ini dapat dipakai untuk sampel yang kecil, sedang dan besar. Jika sampel lebih dari 100 maka dianjurkan mengelompokkan waktu ketahanan dalam beberapa interval.

Taksiran fungsi survival S(t) adalah proses taksiran dari sampel yang bertahan hidup lebih dari waktu yaitu :

$$\hat{S}(t) = \frac{\text{Jumlah individu yang bertahan hidup lebih dari t}}{\text{Jumlah seluruh individu dalam sampel}} \qquad \text{...........(2.6)}$$

Bila diketahui semua individu mati, tetapi dalam pengamatan ada beberapa individu yang bertahan hidup sampai akhir study maka untuk memudahkan dalam perhitungan misalkan n adalah jumlah seluruh individu yang diamati

$t_{(1)} \le t_{(2)} \le \ldots\ldots \le t_{(n)}$

Maka taksiran Produk limit adalah :

$$\hat{S}(t) = \prod_{t(r) \le t} \frac{n - r}{n - r + 1} \qquad \text{...........(2.7)}$$

dimana r berjalan dari bilangan integer positif dengan $t_{(r)} < t$ dan $t_{(r)}$ adalah data tersensor.

Bila terdapat beberapa pengamatan yang sama (tie) maka digunakan harga S (t) yang terkecil.

Secara praktis penaksiran dengan produk-limit untuk data tersensor dapat dihasilkan dengan cara sebagai berikut :

1. Kolom 1 adalah semua waktu ketahanan yang telah diurutkan dari yang terkecil sampai terbesar baik data yang lengkap maupun tersensor. Untuk data tersensor diberi tanda +
2. Kolom kedua berisi rank dari data kolom pertama yang dinotasikan i
3. Kolom ketiga hanyalah berisi rank dari data yang lengkap yang dinotasikan r, dimana r = i untuk data lengkap.
4. Kolom ke-empat berisi nilai (n - r) / (n - r + 1) untuk data lengkap.
5. Kolom kelima berisi nilai dari perkalian (n-r)/(n-r+1) dengan nilai di baris sebelumnya, nilai ini adalah $\hat{S}(t)$

Tabel : 1.2 Contoh Estimasi Survival Analysis dengan produk - limit

Waktu	Rank			
t	i	r	(n - r)/(n - r + 1)	$\hat{S}$ (t)
3	1	1	3/4	3/4
7+	2	-	-	-
9	3	3	1/2	3/4.1/2=3/8
10	4	4	0	0

Digunakan Produk limit dari Kaplan dan Meier untuk mengetahui perbedaan dua distribusi ketahanan hidup secara grafis. Gambaran ketahanan hidup dari produk limit Kaplan Meier ini hanya menunjukkan perbedaan secara kasar saja. Untuk mendapatkan hasil yang lebih akurat diperlukan suatu uji hipotesis.