Model Survival Menggunakan Sebaran Weibull

Julio Adisantoso, G16109011/STK

8 Juli 2010

Ringkasan

Analisis survival merupakan alat statistik yang tujuan utamanya adalah men- ganalisis data yang selalu positif dalam skala pengukuran dengan jarak interval data awal dan akhir yang panjang. Metode analisis survival yang menghubungkan antara waktu sur- vival dengan variabel lain adalah model hazard proporsional dimana formulanya memungkinkan untuk interpretasi pengaruh dari masing-masing variabel bebas akan lebih mudah. Model untuk waktu survival Y dapat menggunakan sebaran exponential, Weibull, gamma, logistic, normal, dan lainnya. Makalah ini membahas model untuk data waktu survival dengan menggunakan sebaran Weibull dibandingkan dengan model sebaran lainnya.

Hasil analisis menunjukkan bahwa kelompok *Treatment* atau kelompok pasien leukemia yang diberi perlakuan memiliki peluang hidup lebih tinggi dibanding kelompok *Control* atau kelompok yang tidak diberi perlakuan khusus. Model untuk waktu survival data yang diberikan dapat menggunakan sebaran exponential, Weibull, atau Lognormal. Namun demikian, model Weibull memiliki statistik AIC dan SBC yang paling kecil sehingga model yang paling sesuai untuk analisis survival terhadap data yang dicobakan adalah model Weibull.

1 Pendahuluan

Salah satu tipe data yang mungkin adalah waktu dari suatu titik awal tertentu sampai muncul kejadian 'gagal'. Data waktu sampai muncul kejadian gagal disebut sebagai **waktu survival**, yang memiliki dua karakteristik, yaitu (a) tidak negatif dan memiliki sebaran dengan ekor yang panjang, dan (b) beberapa subyek mungkin memiliki periode sehingga waktu kejadian gagal tidak diketahui atau secara umum waktu survival yang tidak diketahui. Data yang memiliki karakteristik kedua tersebut dinamakan **tersensor** (Dobson, 2001). Metode analisis statistik pada umumnya akan menghasilkan interpretasi yang bias jika terdapat data yang tidak lengkap atau tersensor.

Analisis survival merupakan alat statistik yang tujuan utamanya adalah menganalisis data yang selalu positif dalam skala pengukuran dengan jarak interval data awal dan akhir yang panjang (McCullagh & Nelder, 1983). Data dengan karakteristik tidak lengkap atau tersensor dan fokus pada pendugaan parameter populasi *life data*, sehingga analisis yang digunakan adalah *life data analysis* (Nelson, 1982). Metode analisis survival yang menghubungkan antara waktu survival dengan variabel lain adalah model *hazard proporsional* dimana formulanya memungkinkan untuk interpretasi pengaruh dari masing-masing variabel bebas akan lebih mudah.

Model untuk waktu survival Y dapat menggunakan sebaran *exponential*, *Weibull*, *gamma*, *logistic*, *normal*, dan lainnya. Makalah ini membahas model untuk data waktu survival dengan menggunakan sebaran *Weibull* menggunakan data pada buku Dobson (2001) Bab 10.

2 Analisis Survival

Analisis survival adalah analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu obyek sampai terjadinya peristiwa gagal (*failure event*). Dalam menentukan waktu survival, Y, terdapat tiga elemen yang harus diperhatikan yaitu waktu awal (*time origin*), definisi *failure time* yang harus jelas, dan skala waktu sebagai satuan pengukuran. Perbedaan antara analisis survival dengan analisis statistik lainnya adalah adanya data tersensor. Menurut Pyke & Thompson (1986) data dikatakan tersensor jika pengamatan waktu survival hanya sebagian, tidak sampai *failure event*. Penyebab terjadinya data tersensor antara lain:

1. *Loss to follow up*, terjadi bila obyek pindah, meninggal atau menolak untuk berpartisipasi.
2. *Drop out*, terjadi bila perlakuan dihentikan karena alasan tertentu.
3. *Termination*, terjadi bila masa penelitian berakhir sementara obyek yang diobservasi belum mencapai *failure event*.

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.

Fungsi survival, $S(y)$, didefinisikan sebagai peluang suatu obyek bertahan setelah waktu ke-y, yaitu

$$S(y) = P(Y \geq y) = 1 - F(y) \tag{1}$$

Fungsi hazard merupakan laju kegagalan sesaat antara selang waktu yang sempit y dan $(y + \delta y)$ dengan asumsi obyek telah bertahan sampai waktu ke-y, yang didefinisikan sebagai

$$\begin{aligned} h(y) &= \lim_{\delta y \to 0} \frac{P(y \leq Y < y + \delta y \mid Y > y)}{\delta y} \\ &= \lim_{\delta y \to 0} \frac{F(y + \delta y) - F(y)}{\delta y} \text{ x } \frac{1}{S(y)} \end{aligned}$$

Karena $f_Y(y)$ merupakan turunan pertama dari $F_Y(y)$ atau

$$\lim_{\delta y \to 0} \frac{F(y+\delta y) - F(y)}{\delta y} = f(y)$$

maka diperoleh

$$h(y) = \frac{f(y)}{S(y)} \tag{2}$$

Dari fungsi survival pada persamaan 1 diperoleh $F(y) = 1 - S(y)$ yang dapat dituliskan sebagai

$$\int f(y)dy = 1 - S(y)$$

dan jika diturunkan terhadap y maka diperoleh

$$f(y) = \frac{d(1 - S(y))}{dy} = -\frac{d}{dy}S(y)$$

Dengan demikian,

$$h(y) = \frac{-\frac{d}{dy}S(y)}{S(y)} \Leftrightarrow -h(y)dy = \frac{d(S(y))}{S(y)}$$

Dengan mengintegralkan $h(y)$ diperoleh

$$\begin{aligned} -\int_0^y h(t)dt &= \int_0^y \frac{1}{S(t)} d(S(t)) \\ -H(y) &= \log[S(y)] \end{aligned}$$

atau

$$H(y) = -\log[S(y)] \tag{3}$$

yang disebut sebagai **fungsi kumulatif hazard**.

Nilai tengah waktu survival umumnya diduga dengan median dari sebaran karena karakteristik data yang memiliki kemiringan. Median waktu survival, $y(50)$, diperoleh dari jawaban persamaan $F(y) = \frac{1}{2}$.

2.1 Sebaran Weibull

Fungsi kepekatan peluang peluang dari sebaran Weibull adalah

$$f(y; \lambda, \theta) = \frac{\lambda y^{\lambda-1}}{\theta^\lambda} \exp\left\{ -\left(\frac{y}{\theta}\right)^\lambda \right\}, \, y \geq 0, \, \lambda > 0, \, \theta > 0$$

Parameter λ dan θ menentukan bentuk dan skala sebaran. Jika $\theta^{-\lambda} = \phi$ maka fungsi kepekatan peluang sebaran Weibull menjadi

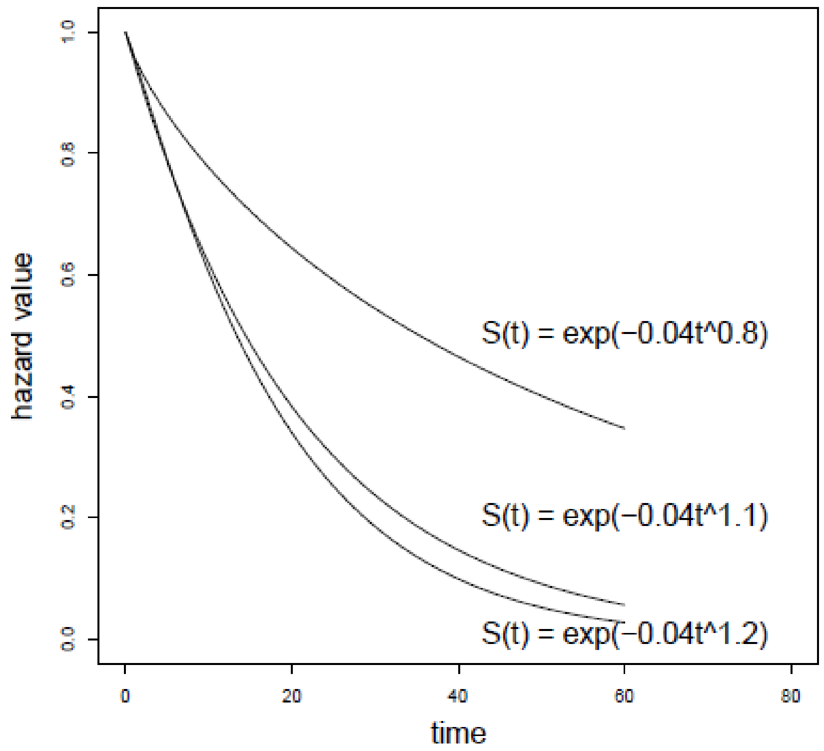
$$f(y; \lambda, \phi) = \lambda \phi y^{\lambda-1} \exp\left(-\phi y^\lambda\right) \tag{4}$$

Jika $\lambda = 1$ maka fungsi sebaran pada persamaan 4 disebut dengan sebaran *exponential*.

Fungsi survival untuk sebaran Weibull adalah

$$\begin{aligned} S(y;\lambda,\phi) &= \int_y^\infty \lambda\phi u^{\lambda-1}\exp\left(-\phi u^\lambda\right)du \\ &= \exp\left(-\phi y^\lambda\right) \end{aligned} \tag{5}$$

Gambar 1 menunjukkan fungsi survival Weibull $S(y)$dengan tiga nilai parameter yang berbeda, yaitu $\phi = 0.04$ dan $\lambda = \{0.8, 1.1, 1.2\}$.



Gambar 1: Fungsi survival Weibull dengan tiga nilai parameter berbeda

Fungsi hazard untuk sebaran Weibull adalah

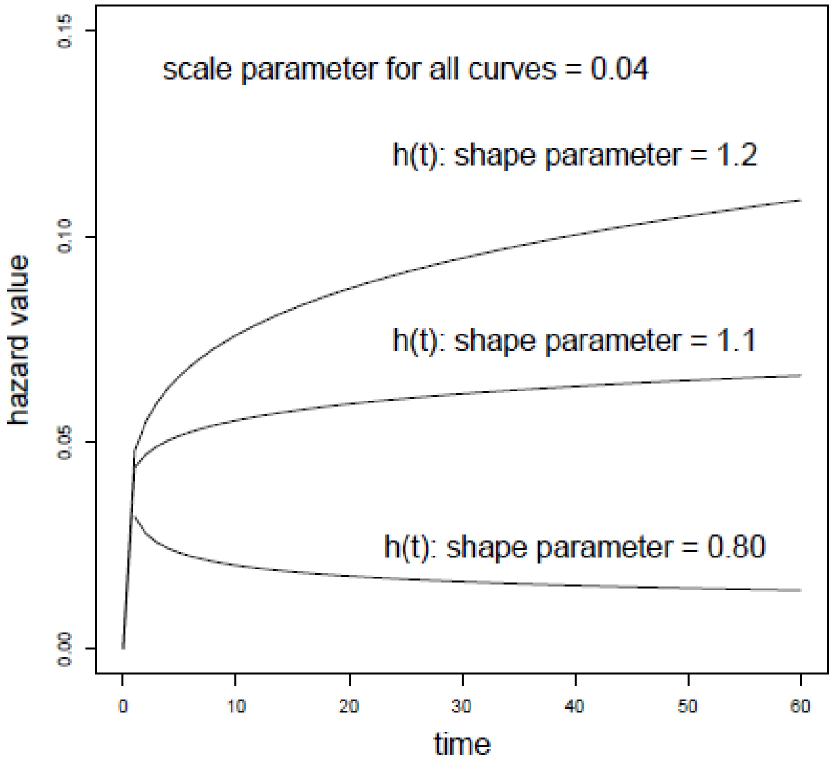
$$h(y;\lambda,\phi) = \lambda\phi y^{\lambda-1} \tag{6}$$

dan fungsi kumulatif hazard adalah

$$H(y;\lambda,\phi) = \phi y^\lambda \tag{7}$$

Gambar 2 menunjukkan fungsi hazard Weibull $h(y)$dengan tiga nilai parameter yang berbeda, yaitu $\phi = 0.04$ dan $\lambda = \{0.8, 1.1, 1.2\}$. Dari persamaan (7) dapat diturunkan fungsi logaritme

$$\log H(y) = \log\phi + \lambda\log y = \log[-\log S(y)] \tag{8}$$



Gambar 2: Fungsi hazard Weibull dengan tiga nilai parameter berbeda

2.2 Model Hazard Proporsional

Jika resiko gagal pada waktu tertentu bergantung pada nilai x_1 ... x_p dari p variabel kovariat, X_1 ... X_p, maka nilai variabel tersebut diasumsikan telah tercatat sebagai *time origin*. Misalkan $h_0(y)$ sebagai fungsi hazard untuk setiap obyek dengan nilai dari semua variabel X adalah nol maka fungsi $h_0(y)$ dikatakan sebagai fungsi *baseline hazard* (Shuo-Jye Wu, 2002). **Model hazard proporsional** atau lebih dikenal dengan **regresi cox** adalah

$$h_1(y;\beta) = h_0(y)\ \exp\left(\sum_{i=1}^{p} x_i\beta_i\right) \tag{9}$$

dan fungsi kumulatif hazard diberikan oleh

$$H_1(y) = \int_0^y h_1(t)dt = \int_0^y h_0(t)e^{\mathbf{x}^T\beta}dt = H_0(y)e^{\mathbf{x}^T\beta}$$

sehingga

$$\log H_1(y) = \log H_0(y) + \sum_{i=1}^{p} x_i\beta_i \tag{10}$$

Fungsi hazard proporsional pada persamaan (10) menunjukkan suatu model linier terampat dengan *link function* **log**. McCullagh & Nelder (1989) menunjukkan bahwa ada beberapa *link function* yang umum digunakan, tergantung pada asumsi sebaran variabel respon y. Jika sebaran y adalah keluarga eksponensial seperti Normal, Gamma, Inverse Normal, dan Poisson maka *link function* yang dapat digunakan antara lain adalah:

- *Identity link* : $f(z) = z$
- *Log link* : $f(z) = log(z)$
- *Power link* : $f(z) = z^a$ untuk nilai tertentu.

Sedangkan jika diasumsikan sebaran y adalah binomial atau multinomial maka dapat digunakan *link function*

- *Logit link* : $f(z) = \log(z/(1-z))$
- *Probit link* : $f(z) = \phi^{-1}(z)$
- *Complementary log-log link* : $f(z) = \log(-\log(1-z))$
- *Log-log link* : $f(z) = -\log(-\log(z))$

Jika variabel kovariat x bernilai biner, yaitu $x_k = 0$ untuk tanpa perlakuan dan $x_k = 1$ untuk perlakuan, maka **rasio Hazard** atau **Hazard relatif** antara ada dan tidak ada perlakuan adalah

$$\frac{h_1(y;\beta)}{h_0(y;\beta)} = e^{\beta_k}$$

menunjukkan bahwa nilai $\sum_{i \neq k} x_i \beta_i$ adalah konstan. Bentuk umum dari model tersebut seperti yang dituliskan pada persamaan (9).

2.3 Fungsi Survival Empirik

Fungsi survival empirik adalah penduga dari peluang survival lebih dari y, diberikan oleh persamaan

$$\widehat{S}(y) = \frac{\text{banyaknya subyek dengan waktu survival} \geq y}{\text{total banyaknya subyek}}$$

Cara yang umum untuk menghitung fungsi survival empirik tersebut menggunakan penduga **Kaplan Meier**, yang juga disebut sebagai penduga **product limit**. Pertama dilakukan pengurutan secara menaik dari data waktu survival sehingga $y_{(1)} \leq y_{(2)} \leq ... \leq y_{(k)}$. Jika n_j melambangkan banyaknya subyek yang hidup sebelum $y_{(j)}$ dan d_j melambangkan banyaknya kematian terjadi selama selang waktu yang kecil, $y_{(j)} - \delta$ sampai $y_{(j)}$, maka penduga Kaplan Meier untuk fungsi survival pada waktu y adalah

$$\widehat{S}(y) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right) \tag{11}$$

untuk y antara $y_{(j)}$ dan $y_{(j+1)}$.

Fungsi survival empirik $\widehat{S}(y)$ dapat diplot antara log-log pada persamaan (8) dengan logaritme dari waktu survival, $\log(y)$. Untuk sebaran Weibull, kesesuaian ditunjukkan adanya garis yang relatif cukup lurus daris plot tersebut.

2.4 Pendugaan

Untuk subyek ke-j, data yang dicatat meliputi: waktu survival y_j, indikator sensor: $\delta_j = 1$ jika waktu survival tidak tersensor dan $\delta_j = 0$ jika tersensor, serta vektor variabel kovariat $\mathbf{x}_j$. Misalkan y_1,...,y_r melambangkan pengamatan yang tidak tersensor, dan y_{r+1},...,y_n melambangkan pengamatan yang tersensor, maka fungsi log-likelihood adalah

$$
\begin{aligned}
l &= \sum_{j=1}^{n} \{\delta_j \log f(y_j) + (1-\delta_j) \log S(y_j)\} \\
&= \sum_{j=1}^{n} \{\delta_j \log h(y_j) + \log S(y_j)\} \qquad (12)
\end{aligned}
$$

Jika data untuk subyek j adalah $\{y_j, \delta_j, \text{dan } \mathbf{x}_j\}$ dan model menggunakan sebaran Weibull, maka fungsi log-likelihood adalah

$$
l = \sum_{j=1}^{n} \left\{\delta_j \log(\lambda \alpha y_j^{\lambda-1} e^{\mathbf{x}^T\beta}) - (\lambda y_j^{\lambda} e^{\mathbf{x}^T\beta})\right\} \qquad (13)
$$

3 Analisis Data

3.1 Bahan dan Metode

Analisis data dilakukan terhadap contoh Tabel 10.1 pada buku Dobson (2001), yaitu waktu hidup pasien leukemia. Subyek terbagi menjadi dua kelompok masing-masing 21 pasien, yaitu kelompok kontrol yang diberi plasebo dan kelompok perlakuan yang diberi 6 *mercaptopurine* (Tabel 1).

Tabel 1: Data yang dianalisis

textitControl										
1	1	2	2	3	4	4	5	5	8	8
8	8	11	11	12	12	15	17	22	23	
Treatment: 6 *mercaptopurine*										
6	6	6	6*	7	9*	10	10*	11*	13	16
17*	19*	20*	22	23	25*	32*	32*	34*	35*	
* menunjukkan sensor										

Model hazard proporsional dengan sebaran Weibull adalah

$$
h(y) = \lambda y^{\lambda-1} \exp(\beta_0 + \beta_1 x),\ y \sim \text{Weibull}
$$

dimana $x = 0$ untuk kelompok *Control*, $x = 1$ untuk kelompok *Treatment*, dan λ adalah parameter bentuk sebaran Weibull.

3.2 Hasil Analisis

Data disusun dan dibaca dengan prosedur SAS sebagai berikut:

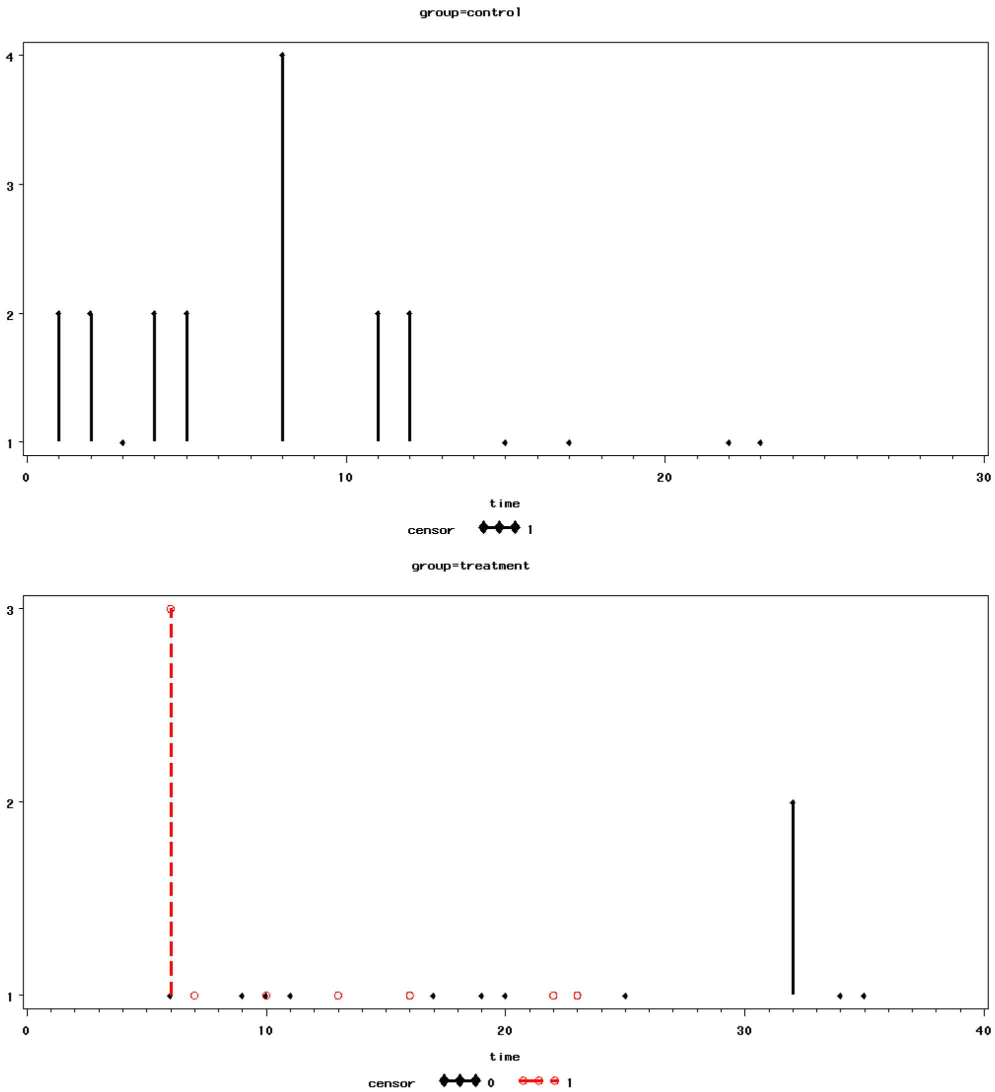
```
data table10_1;
length group $10;
input group $ censor time;
datalines;
control   1  1
control   1  1
...
treatment 0 35;
run;

data table10_1b;
set table10_1;
if group = 'treatment' then gp = 1; else gp = 0;
run;
```

Gambar 3 menunjukkan plot data waktu hidup pasien leukemia untuk tiap kelompok dimana simbol titik menunjukkan data tidak tersensor dan simbol lingkaran menunjukkan data tersensor. Pada kelompok data *Control* tidak terdapat data tersensor yang berbeda dengan kelompok data *Treatment* dimana terdapat 12 data tersensor. Berdasarkan Gambar 4 dapat dilihat bahwa data memiliki sebaran yang miring (*skew*) dan menunjukkan bahwa waktu hidup lebih panjang untuk kelompok *Treatment* (garis putus-putus). Disamping itu, kelompok *Treatment* memiliki peluang hidup lebih tinggi dibanding kelompok *Control*. Hal ini juga dapat dilihat pada Tabel 2.

Tabel 2: Nilai-nilai penduga Kaplan Meier dari fungsi survivor

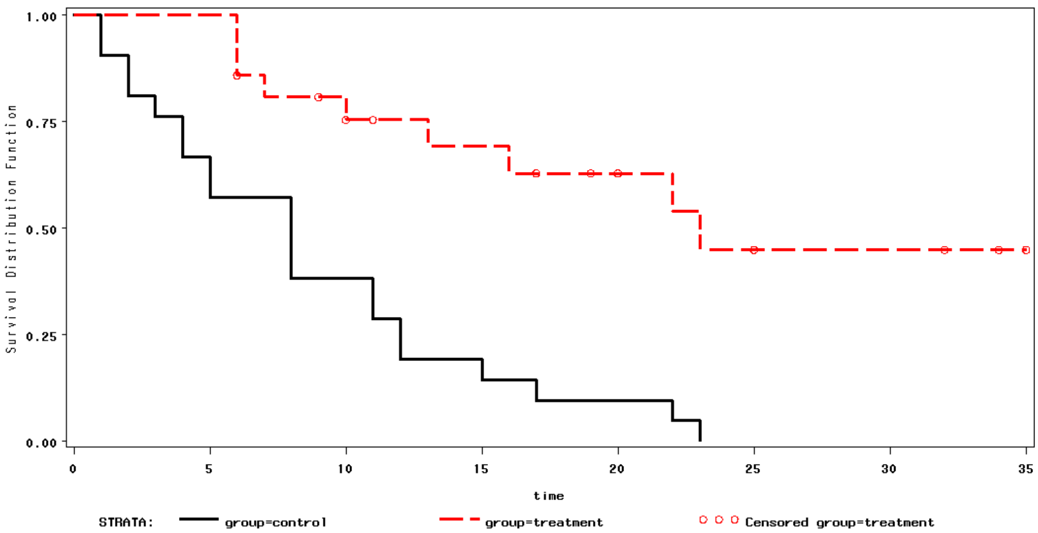
Control				*Treatment*			
Waktu (y_j)	n_j	d_j	$\widehat{S}(y)$	Waktu (y_j)	n_j	d_j	$\widehat{S}(y)$
0-<1	21	0	1.0000	0-<6	21	0	1.0000
1-<2	21	1	0.9048	6-<7	21	3	0.8571
2-<3	19	1	0.8095	7-<10	17	1	0.8067
3-<4	17	0	0.7619	10-<13	15	1	0.7529
4-<5	16	1	0.6667	13-<16	12	1	0.6902
5-<8	14	1	0.5714	16-<22	11	1	0.6275
8-<11	12	3	0.3810	22-<23	7	1	0.5378
11-<12	8	1	0.2857	≥23	6	1	0.4482
12-<15	6	1	0.1905				
15-<17	4	0	0.1429				
17-<22	3	0	0.0952				
22-<23	2	0	0.0476				
≥23	1	0	0.0000				



Gambar 3: Plot data waktu hidup pasien leukemia untuk tiap kelompok

Gambar 5 menunjukkan plot logaritme dari fungsi kumulatif Hazard ($\log H_1(y)$) dengan $\log y$, dimana garis putus-putus adalah kelompok *Treatment* sedangkan garis utuh adalah kelompok *Control*. Kedua garis terlihat cukup lurus yang menunjukkan bahwa sebaran Weibull adalah sesuai (8). Disamping itu, kedua garis terlihat sejajar yang menunjukkan bahwa model proporsional Hazard juga sesuai. Kemiringan garis bernilai mendekati satu karena membentuk sudut sekitar 45^o berarti bahwa sebaran eksponensial sama baiknya dengan sebaran Weibull ($\lambda = 1$). Jarak antar kedua garis sekitar 1.4 yang menunjukkan bahwa rasio Hazard kira-kira sebesar $\exp(1.4) \approx 4$.

Pengamatan terhadap gambar atau grafik sering menimbulkan kesulitan dalam penentuan kesesuaian sebaran karena pengaruh skala. Disamping itu, pola bebera-



Gambar 4: Grafik penduga *Kaplan Meier* waktu hidup pasien untuk tiap kelompok

Tabel 3: Nilai-nilai statistik kesesuaian model

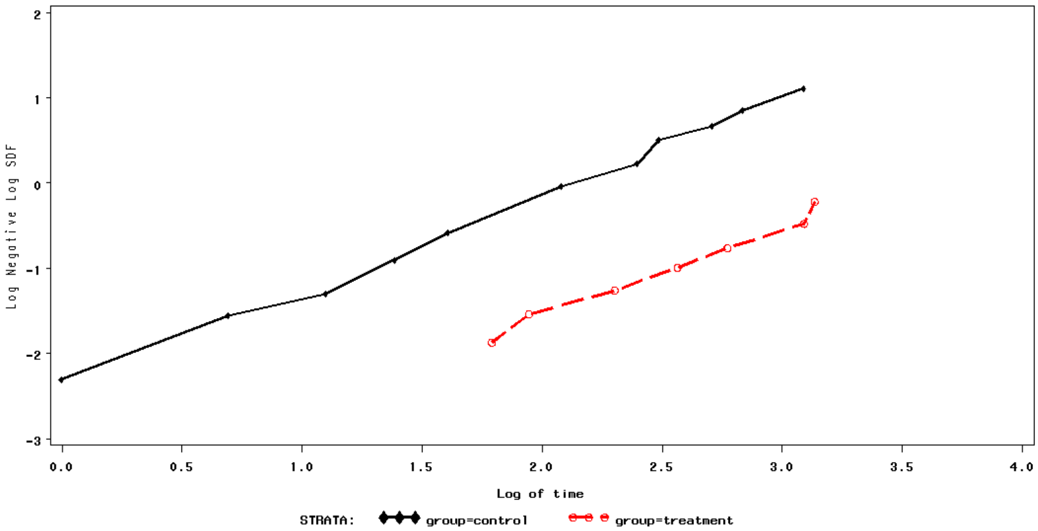
Model	Maximum Log-likelihood	AIC	SBC
WEIBULL	-47.064	100.128	105.341
NORMAL	-117.968	241.935	247.148
LNORMAL	-47.189	100.378	105.591
LOGISTIC	-118.459	242.917	248.130
LLOGISTIC	-48.146	102.292	107.505
EXPONENTIAL	-49.009	102.017	105.493
GAMMA	-46.874	101.748	108.699
LOG GAMMA	-110.033	228.066	235.016

pa garis di dalam grafik sering kali sulit dibedakan. Oleh karena itu, pengamatan terhadap grafik harus diperkuat dengan nilai statistik tertentu. Beberapa nilai statistik yang dapat digunakan untuk menentukan model yang sesuai antara lain adalah AIC (*Akaike Information Criteria*) dan SBC (*Schwarz's Bayesian Criterion*) dengan formula

$$-2 * (\text{Log-Likelihood}) + k * (p)$$

dimana p adalah banyaknya paremeter, dan $k = 2$ untuk AIC atau $k = \log(n)$, n adalah banyaknya pengamatan, untuk SBC.

Tabel 3 menunjukkan nilai-nilai statistik yang dapat digunakan untuk menentukan model yang sesuai, sedangkan Lampiran 1 mencantumkan program macro SAS yang digunakan untuk menganalisis data dan menghitung nilai AIC dan SBC untuk beberapa sebaran. Diantara semua sebaran yang dicobakan, model Weibull, Log Normal, dan Eksponential yang memiliki nilai AIC dan SBC kecil.



Gambar 5: Fungsi log kumulatif Hazard

Tabel 4: Penduga parameter model Weibull

Parameter	Estimate	Standard Error	Asymptotic Normal 95% Confidence Limits Lower	Upper
Intercept	2.2484	0.1660	1.9231	2.5737
gp	1.2673	0.3106	0.6585	1.8762
EV Scale	0.7322	0.1078	0.5486	0.9772
Weibull Shape	1.3658	0.2012	1.0233	1.8228

Namun demikian, model Weibull memiliki statistik AIC dan SBC yang paling kecil. Oleh karena itu, model yang sesuai untuk analisis survival terhadap data yang dicobakan adalah model Weibull.

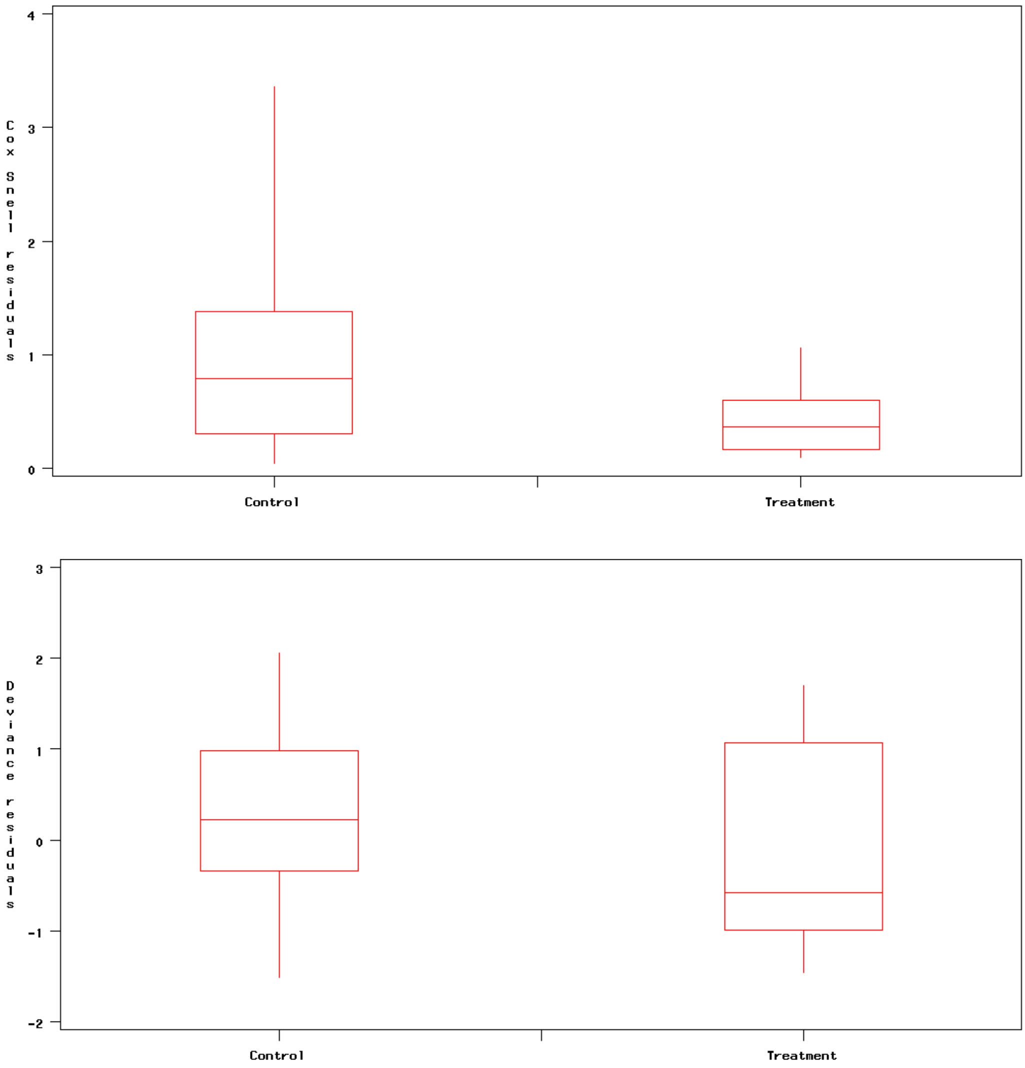
Untuk model yang sesuai, yaitu Weibull, diperoleh penduga parameter model seperti tercantum pada Tabel 4. Dengan demikian, model Hazard proporsional menjadi

$$h(y) = 1.3658y^{0.3658}e^{2.2484+1.2673x}$$

Untuk melihat apakah sebaran atau model Exponential sesuai untuk data tersebut dibanding model Weibull, dapat dilakukan dengan melakukan pengujian hipotesis bahwa $\lambda = 1$ dengan menggunakan statistik Wald, yaitu

$$z = \frac{1.3658 - 1}{0.2012} = 1.8181$$

atau menggunakan statistik Deviance $D = 2(l_w - l_e)$, l adalah *maximum log-likelihood*, sehingga $D = 2(-47.064 + 49.009) = 3.89$. Jika nilai z dibandingkan dengan nilai sebaran normal, atau nilai D dibandingkan dengan sebaran χ^2 maka terdapat kecenderungan untuk menerima hipotesis bahwa $\lambda = 1$ atau sebaran Exponential juga sesuai digunakan di dalam model. Namun demikian, berdasarkan selang kepercayaan 95% dari λ pada Tabel 4 yaitu 1.0233< λ <1.8228 maka terdapat kecenderungan bahwa nilai $\lambda > 1$ yang artinya model Weibull lebih sesuai.



Gambar 6: Plot sisaan *Cox-Snell* dan *Deviance* model Weibull

Gambar 6 adalah plot sisaan *Cox-Snell* dan *Deviance* untuk model Weibull, yang menunjukkan bahwa sisaan *Cox-Snell* simetris dibanding sisaan *Deviance*. Disamping itu, perbedaan pola sebaran dari kelompok data *Control* dan *Treatment* menunjukkan bahwa model belum dapat menjelaskan sepenuhnya pola survival dari kedua kelompok data tersebut.

4 Kesimpulan

Berdasarkan analisis yang telah dilakukan telah diperoleh beberapa kesimpulan, yaitu model untuk waktu survival data yang diberikan dapat menggunakan sebaran exponential, Weibull, atau Lognormal. Model ini memberikan nilai statistik AIC dan SBC yang kecil dibanding sebaran lainnya yang dicobakan. Namun demikian, model Weibull memiliki statistik AIC dan SBC yang paling kecil sehingga model yang paling sesuai untuk analisis survival terhadap data yang dicobakan adalah model Weibull.

Model regresi proporsional Hazard menghubungkan pengaruh kovariat dengan menggunakan *link function* log. *Link function* yang dapat digunakan antara lain adalah *identity* atau *power*.

Berdasarkan plot data waktu hidup pasien leukemia untuk tiap kelompok menunjukkan bahwa kelompok *Treatment* atau kelompok yang diberi perlakuan memiliki peluang hidup lebih tinggi dibanding kelompok *Control* atau kelompok yang tidak diberi perlakuan khusus.

5 Daftar Pustaka

Agresti, A. 2007. ***An Introduction to Categorical Data Analysis***. 2^{nd} Ed. John Wiley and Sons, Inc.

Dobson, A.J. 2001. ***An Introduction to Generalized Linear Models***. Chapman Hall/CRC Texts in Statistical Science Series.

Gharibvand, L; D.R.Jeske, & S.Liao. 2000. ***Evaluation of a Hospice Care Referral Program Using Cox Proportional Hazards Model***. SAS Institute Inc.

McCullagh,P. and Nelder,J.A. 1983. ***Generalized Linear Models***. 2^{nd} Ed. Chapman and Hall.

Southey, B.R.; S.L.Rodriguez-Zas; & K.A.Leymaster. 2003. ***Discrete Time Survival Analysis of Lamb Mortality in a Terminal Sire Composite Population***. Journal of Animal Sciences 2003. 81:1399-1405.

Wei-Wang. 2004. ***Proportional Hazards Regression Models with Unknown Link Function and Time-Dependent Covariates***. Statistica Sinica 14(2004), 885-905. Harvard University.

6 Lampiran

Lampiran 1. Program Macro SAS untuk menghitung statistik kesesuaian model

```
%macro AIC_SBC(dataset=,time=,censor=,covariates=,mi=50);
   * mi=maksimum iterasi, default 50;
   %if %length(&mi) = 0 %then %let mi=50;

   * PROC LIFEREG tanpa option NOLOG;
   %LET distr = exponential weibull gamma normal lnormal
                logistic llogistic;

   %do i=1 %to 7;
       proc lifereg data=&dataset outest=_%scan(&distr,&i)_;
            %scan(&distr,&i): model &time*&censor(0)=&covariates /
            dist=%scan(&distr,&i) maxiter=&mi;
            title "Macro AIC_SBC";
       run;
   %end;

   * PROC LIFEREG dengan option NOLOG;
   proc lifereg data=&dataset outest=_EXPONENTIAL_NL_;
      one_param_EXTREME_VALUE: model &time*&censor(0)=
      &covariates / dist=exponential NOLOG maxiter=&mi;
   run;
   proc lifereg data=&dataset outest=_WEIBULL_NL_;
      two_param_EXTREME_VALUE: model &time*&censor(0)=
      &covariates / dist=weibull NOLOG maxiter=&mi;
   run;
   proc lifereg data=&dataset outest=_GAMMA_NL_;
      LOG_GAMMA: model &time*&censor(0)=
      &covariates / dist=gamma NOLOG maxiter=&mi;
   run;

   * hitung banyaknya kovariat;
   %LET nvar=0;
   %do %while (%length(%scan(&covariates,&nvar+1))>0);
      %LET nvar=%eval(&nvar+1);
   %end;

   * hitung banyaknya pengamatan;
   proc sql noprint;
      select count(*) into :n
      from &dataset
         %if %length(&covariates) = 0 %then
      where &time>.; %else %do;
      where
         nmiss(%do i=1 %to &nvar; %scan(&covariates,&i),
            %end;
            &time)=0
      %end;;
      quit;

   * hitung AIC dan SBC tiap model;
   data _models_;
   set _weibull_ _normal_ _lnormal_ _logistic_ _llogistic_
       _exponential_ (in=e) _gamma_ (in=g)
       _weibull_nl_ _exponential_nl_ (in=enl) _gamma_nl_ (in=gnl);
       if e or enl then k=1; else if g or gnl then k=3; else k=2;
       k=k+&nvar;
       AIC=-2*_LNLIKE_ + 2*k;
       SBC=-2*_LNLIKE_ + log(&n)*k;
       label k='Number of parameters';
       drop &time _type_ _name_;
       run;
   * cetak hasil;
   proc print data=_models_ noobs label;
       label _LNLIKE_='Maximum log-likelihood'
       _MODEL_ ='Model';
       var _model_ _LNLIKE_ AIC SBC;
       run;
%mend AIC_SBC;
```