

Analisis Data Survival Menggunakan Metode *Proportional Hazard* dan *Accelerated Failure Time*

Sri Astuti Thamrin*

Abstrak

This paper describes survival data analysis using proportional hazard method and accelerated failure time method to determine the factors that influence survival times of patients' Diabetes Mellitus in Dr. Wahidin Sudirohusodo hospital in Makassar. Besides that, we compare the results of these methods. The results show that among eight tested variables, three of them are affected factors.

Kata Kunci : *survival data, proportional hazard method, accelerated failure time method, survival time, Diabetes Mellitus.*

1. Pendahuluan

Dalam era globalisasi dewasa ini, aplikasi statistika berkembang dengan pesat sebagai suatu bidang kajian terapan. Beberapa metode aplikasi statistika berkembang sebagai bentuk respons atas berbagai masalah yang ada pada berbagai bidang, seperti kedokteran, pertanian, teknik, ekonomi dan sebagainya. Salah satu aplikasi statistik untuk bidang kedokteran diterapkan untuk menganalisis kelangsungan hidup (*Survival Analysis*) pasien. Analisis kelangsungan hidup bertujuan memodelkan distribusi yang mendasari variabel waktu kegagalan (*failure time*) dan untuk menaksir ketergantungan variabel waktu kegagalan dengan variabel bebas. Menurut Galbriath (2004), analisis kelangsungan hidup adalah analisis data yang menggambarkan waktu terjadinya sebuah kejadian tertentu ke titik akhir kejadian itu ("*time to event*" data). Contoh untuk titik akhir sebuah kejadian misalnya kematian, sembuh dari suatu penyakit, serta munculnya kembali gejala suatu penyakit.

Data kelangsungan hidup terdiri dari variabel respon dan variabel bebas. Variabel respon menggambarkan lamanya waktu sampai terjadi kejadian spesifik (waktu kejadian/ *event time*, waktu kegagalan/ *failure time* atau waktu kelangsungan hidup). Variabel bebas dapat berupa diskrit seperti jenis kelamin, ras, atau kontinu seperti umur, temperatur.

Karakteristik dari data kelangsungan hidup yaitu adanya pengamatan tersensor. Contohnya dalam bidang kedokteran sensor ini timbul dari adanya pasien yang mengundurkan diri dari penelitian atau penelitian telah selesai sebelum adanya titik akhir dari kejadian (Galbriath, 2004). Titik akhir dari kejadian misalnya kematian, sembuh dari suatu penyakit, serta munculnya kembali gejala suatu penyakit. Adanya sensor pada analisis kelangsungan hidup merupakan salah satu hal yang membedakan dengan analisis statistik lainnya. Terdapat 2 alasan utama mengapa data kelangsungan hidup tidak dapat menggunakan model regresi berganda yaitu

- 1) Variabel dependen tidak berdistribusi normal yaitu waktu kelangsungan hidup (*survival time*) biasanya menggunakan distribusi eksponensial, distribusi weibull, distribusi gamma, distribusi lognormal atau distribusi log-logistik.
- 2) Adanya data sensor yang mengakibatkan beberapa observasi tidak lengkap.

* Staf pengajar pada Jurusan Matematika FMIPA Universitas Hasanuddin Makassar

Hal yang penting dalam analisis kelangsungan hidup adalah menentukan berapa besar peluang dapat bertahan hidup suatu individu sampai terjadinya suatu kejadian yang spesifik, serta menentukan faktor-faktor yang mempengaruhi kelangsungan hidup objek/individu. Untuk itu ada beberapa metode pendugaan yang dapat dilakukan, baik secara parameter maupun non-parameter. Pendugaan secara parameter dilakukan apabila objek/individu diamati dan diukur pada interval waktu yang sama, diantaranya dilakukan dengan metode regresi *Cox Proportional Hazard*, regresi logistik secara parsial and *Accelerated Failure Time*. Sedangkan pendugaan secara non-parameter dilakukan apabila objek/individu diamati dan diukur dalam interval waktu yang tidak sama, dapat dilakukan dengan menggunakan metode *Aktuarial* dan *Kaplan-Meier*.

Berdasarkan hal-hal tersebut diatas, dalam penelitian ini akan dibahas model *accelerated failure time* dan model *proportional hazards* dalam menentukan faktor-faktor yang mempengaruhi kelangsungan hidup penderita Diabetes Mellitus (DM) pada Rumah Sakit Wahidin Sudirohusodo, dari Januari 2005 sampai Mei 2006, serta membandingkan kedua model tersebut.

2. Tinjauan Pustaka

2.1. Gambaran Distribusi Waktu Antar Kejadian

Fungsi distribusi kumulatif dari variabel acak T didefinisikan sebagai

$$F(t) = P[T \leq t], t \geq 0,$$

dimana T adalah waktu kelangsungan hidup, $F(t)$ adalah peluang acak dari subjek yang terpilih dari populasi yang akan meninggal sebelum waktu t . Jika T adalah variabel acak kontinu dengan fungsi kepadatan $f(t)$ maka hubungannya dengan $F(t)$ adalah

$$f(t) = \frac{dF(t)}{dt}, F(t) = \int_0^t f(u)du.$$

Oleh karena itu, fungsi kelangsungan hidup sebagai peluang individu akan bertahan hidup sampai waktu t adalah

$$S(t) = P[T \geq t] = 1 - F(t).$$

Dalam analisis kelangsungan hidup dikenal pula adanya *hazards rate*. *Hazards rate* didefinisikan sebagai limit/batas dari angka kematian jika interval waktu diperoleh sangat kecil, yaitu

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{P[t \leq T < t + h | T \geq t]}{h}.$$

Sementara itu, angka kematian dalam waktu t pada beberapa unit waktu didefinisikan sebagai peluang individu meninggal antara waktu t dan $t+1$ yang hidup pada waktu t dituliskan dengan $m(t) = P[t \leq T < t + 1 | T \geq t]$.

2.2. Model *Proportional Hazard*

Model *proportional hazards* mengasumsikan bahwa angka risiko (*hazard rate*) merupakan fungsi dari variabel bebas (*covariate*).

Suatu model regresi dapat didefinisikan oleh fungsi risiko sebagai fungsi dari *covariate* z , yaitu

$$L(t) \equiv L(t | z) \text{ dan } L(t | z) = L_0(t) e^{z_1 \beta_1 + \dots + z_p \beta_p} = L_0(t) e^{z\beta}$$

dimana z adalah *covariate*, β adalah koefisien regresi, dan $L_0(t) = L(t | z = 0)$. Jika T adalah variabel acak berdistribusi Weibull, maka $L_0(t) = \rho \lambda^\rho t^{\rho-1}$ dengan fungsi risiko bersyarat adalah $L(t | z) = \rho \lambda^\rho t^{\rho-1} e^{z\beta}$ dan fungsi kelangsungan hidup dari T bersyarat z adalah

$$S(t | z) = \left[e^{-\int_0^t \lambda_0(u) du} \right]^{\exp(z\beta)}.$$

2.3. Model Accelerated Failure Time

Model *accelerated failure time* biasa juga disebut model log linier. Model ini mengasumsikan bahwa *covariate* berhubungan linear dengan log dari waktu kegagalan.

Model regresi log-linear didefinisikan oleh pemodelan log T adalah

$$\log T_i = Y = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip} + \sigma \varepsilon_i,$$

dimana β_0, \dots, β_p adalah koefisien regresi, z_{i1}, \dots, z_{ip} adalah *covariate*, ε_i adalah gangguan acak dan σ adalah parameter dari error. Jika diasumsikan bahwa distribusi dari T adalah distribusi Weibull maka fungsi kelangsungan hidup dari T adalah

$$S(t | z) = \exp \left\{ - \left[t e^{z\beta} \right]^{\frac{1}{\sigma}} \right\}$$

dan Log fungsi resiko bersyarat adalah

$$\log L(t | z) = \left(\frac{1}{\sigma} - 1 \right) \log t - \log \sigma - z \left(\frac{\beta}{\sigma} \right)$$

3. Metodologi

Subyek penelitian yang digunakan dalam penelitian ini adalah semua penderita penyakit Kencing Manis yang pernah atau sedang menjalani rawat inap di RS Wahidin Sudirohusodo, Makassar. Pada penelitian ini, populasi dibatasi pada penderita penyakit Kencing Manis yang pernah atau sedang menjalani rawat inap di RS Wahidin Sudirohusodo pada periode Januari 2005 s/d Mei 2006. Sampel penelitian diambil sebanyak penderita yang ada pada periode Januari 2005 sampai Mei 2006. Variabel independen yang digunakan dalam penelitian ini adalah sebanyak 8 variabel yang terdiri dari 5 variabel kategorik dan 3 variabel kontinu. Variabel kategorik yaitu jenis kelamin, jenis penyakit, komplikasi, diet DM dan Kategori NIDDM. Variabel kontinu yaitu umur, kadar gula darah dan berat badan. Sedangkan variabel tak bebas yaitu lama dirawat yang diindex berdasarkan status kelangsungan hidup. Tahapan-tahapan penelitian yang dilakukan adalah:

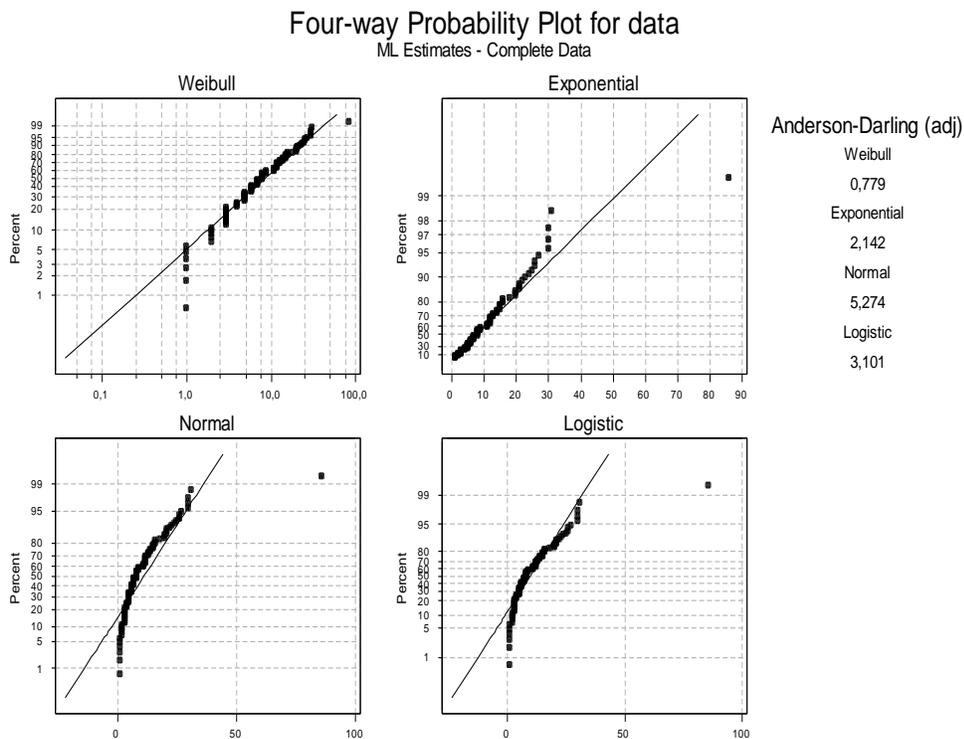
1. Mendeteksi distribusi data untuk melihat ketepatan distribusi dengan data melalui metode maksimum likelihood. Dalam hal ini digunakan empat jenis distribusi data yaitu distribusi weibull, eksponensial, normal, logistik.

2. Pendeteksian awal variabel yang berpengaruh untuk menentukan faktor-faktor yang berpengaruh pada penyakit diabetes mellitus dengan menggunakan uji *equality across strata*. Untuk data bertipe diskrit digunakan uji log-rank, sedangkan untuk data bertipe kontinu digunakan regresi cox. Prediktor akan dimasukkan kedalam model jika nilai p -value $< 0,25$.
3. Menentukan faktor yang berpengaruh pada penyakit diabetes mellitus dengan menggunakan model *proportional hazards* dan model *accelerated failure time*, serta membandingkan hasil yang diperoleh dari kedua metode tersebut.

4. Pembahasan

Data yang digunakan dalam penelitian ini adalah data hasil pencatatan medik pasien penyakit diabetes mellitus (DM) sebanyak 102 orang dari bulan Januari tahun 2001 sampai Mei tahun 2006 di Rumah Sakit Dr. Wahidin Sudirohusodo Makassar. Data yang diperoleh diolah dengan menggunakan software Minitab 13 dan Stata 8,0.

Penentuan distribusi yang mendasari data ditunjukkan dari bentuk plot garis lurus dan nilai Anderson-Darling yang kecil. Dari hasil plot pada gambar di bawah menunjukkan bahwa plot probability yang paling mendekati lurus adalah distribusi weibull dengan nilai Anderson-Darling paling kecil yaitu 0,779. Oleh karena itu akan digunakan distribusi Weibull.



J

Gambar 1. Uji Anderson-Darling untuk Menentukan Distribusi Data

Uji Cox dan Uji log-rank dilakukan sebagai pendeteksian awal variabel yang berpengaruh yang akan dimasukkan ke dalam analisis lanjutan. Berdasarkan uji Cox dan uji log-

rank, maka variabel yang berpengaruh yang akan dimasukkan pada analisis lanjutan adalah gula darah, komplikasi dan diet.

Dengan menggunakan model *proportional hazards* menunjukkan bahwa faktor yang berpengaruh pada penyakit diabetes mellitus adalah diet dan komplikasi dengan dengan interval kepercayaan 95%. Penentuan faktor yang berpengaruh ini dapat dilihat dari nilai p -value $< 0,05$.

Kemudian dengan menggunakan model *accelerated failure time* menunjukkan bahwa faktor yang berpengaruh pada penyakit diabetes mellitus adalah diet dan komplikasi dengan interval kepercayaan 95%. Penentuan faktor yang berpengaruh ini dapat dilihat dari nilai p -value $< 0,05$.

Model terbaik adalah model yang mengandung variabel diet dan komplikasi. Perbandingan kedua model ini dapat dilihat dari nilai hazard ratio pada model *proportional hazards* dan time ratio pada model *accelerated failure time*.

Tabel 1. Hasil Pengolahan STATA 8.

lama	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Ikomp7_2	4.480421	2.58366	2.601	0.009	1.447	13.87296
Idiet5_2	12.27751	6.632259	4.642	0.000	4.258877	35.39368
/ln_p	.005743	.1875374	0.031	0.976	-.3618236	.3733096
p	1.00576	.1886175			.6964052	1.452534
1/p	.9942735	.1864635			.6884521	1.435946

lama	Tm. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Ikomp7_2	.2251184	.138942	-2.416	0.016	.0671512	.7546897
Idiet5_2	.0826278	.0523449	-3.936	0.000	.023872	.2859988
/ln_p	.005743	.1875374	0.031	0.976	-.3618236	.3733096
p	1.00576	.1886175			.6964052	1.452534
1/p	.9942735	.1864635			.6884521	1.435946

Nilai taksiran dari ρ sama untuk kedua model. Tabel berikut ini menunjukkan perbandingan nilai taksiran dan nilai standard error dari model *proportional hazards* dan model *accelerated failure time* dengan menggunakan distribusi weibull.

Tabel 2. Perbandingan Hazard Ratio dan Time Ratio.

Variabel	Koefisien	Hazard ratio	Koefisien	Time ratio
Diet5	taksiran	2,507769	12,27751	0,082627
	se	0,5401956	6,632259	0,052344
Komp7	taksiran	1,499717	4,480421	0,2251184
	se	0,5766557	2,58366	0,138942
ρ	taksiran	1,00576	1,00576	1,00576
	se	0,188617	0,188617	0,188617

Hazard ratio untuk variabel diet5=12,28. ini berarti bahwa risiko kegagalan untuk pasien yang tidak diet adalah 12,28 kali dari risiko kegagalan pasien yang melakukan diet. Time ratio dari diet5 adalah 0,08. Hal ini berarti bahwa rata-rata waktu gagal pasien yang tidak diet adalah 0,08 kali dari pasien yang melakukan diet. Hazard ratio dari variabel komp7 adalah 4,48. Ini berarti bahwa risiko kegagalan untuk pasien yang menderita penyakit kaki diabet dan amputasi diabetes adalah 4,48 kali dari risiko kegagalan pasien penderita disfungsi ereksi, kegagalan ginjal, gangguan mata dan kebutaan, nyeri seluruh tubuh/gangguan persepsi, stroke/lumpuh (hipertensi), gangguan jantung dan saluran cerna, atau tanpa komplikasi. Time ratio dari komp7 adalah 0,23. ini berarti bahwa rata-rata waktu gagal pasien yang menderita komplikasi kaki diabet dan amputasi diabetes adalah 0,23 kali dari pasien penderita disfungsi ereksi, kegagalan ginjal, gangguan mata dan kebutaan, nyeri seluruh tubuh/gangguan persepsi, stroke/lumpuh (hipertensi), gangguan jantung dan saluran cerna, atau tanpa komplikasi.

Selanjutnya *time ratio* dari tidak diet adalah 0,08. Hal ini berarti bahwa rata-rata waktu gagal pasien yang tidak diet adalah 0,08 kali dari pasien yang melakukan diet. Sementara itu time ratio dari komplikasi berupa kaki diabet dan amputasi adalah 0,23, ini menunjukkan bahwa rata-rata waktu kegagalan pasien DM yang menderita komplikasi kaki diabet dan amputasi adalah 0,23 kali dari pasien DM yang menderita komplikasi lainnya.

5. Kesimpulan dan Saran

Dari pembahasan di atas, dapat diambil kesimpulan sebagai berikut:

1. Model *Proportional Hazard* adalah model regresi survival yang mengasumsikan angka resiko (Hazard Ratio) merupakan fungsi dari variabel bebas. Model *Accelerated Failure Time* adalah model yang mengasumsikan bahwa covariate berhubungan linear dengan log dari waktu kegagalan.
2. Pendeteksian variabel yang berpengaruh pada kelangsungan hidup pasien diabetes millitus dengan menggunakan model *Proportional Hazard* dan model *Accelerated Failure Time* memberikan hasil yang sama yaitu faktor komplikasi dan diet.
3. Pada model *Proportional Hazard* pasien yang tidak diet mempunyai risiko gagal yang lebih besar dibandingkan dengan yang melakukan diet dan pasien yang mengidap penyakit komplikasi kaki diabet dan amputasi diabetes mempunyai risiko gagal lebih besar dibanding dengan mengidap komplikasi lainnya atau tidak ada komplikasi.
4. Pada model *Accelerated failure time* yang tidak diet mempunyai rata-rata waktu gagal pasien lebih kecil dibanding dengan yang melakukan diet dan pasien yang mengidap penyakit komplikasi kaki diabet dan amputasi diabetes mempunyai rata-rata waktu gagal lebih kecil dibanding dengan yang mengidap komplikasi lainnya atau tidak ada komplikasi.

Melihat data yang digunakan berdistribusi weibull, sebaiknya dapat diterapkan juga pada data lain yang berdistribusi eksponensial dengan menggunakan metode parametrik: *Proportional Hazard*, *Accelerated Failure Time* dan *Time varying Covariate* untuk data baik berdistribusi Weibull ataupun berdistribusi eksponensial.

Daftar Pustaka

- [1]. Anynomous, 2004, "Parametric and semiparametric methods for survival data", *Module 6, Advance Biostatistics*, The University of Newcastle, Australia.

- [2]. _____, 2004. "Accelerated failure time models and time varying covariates methods for survival data", *Module 7, Advance Biostatistics*, The University of Newcastle, Australia.
- [3]. _____, 2004. "Regression with time to event outcomes", *Lectures Notes 16, Biostatistics*, <http://courses.washington.edu/b515/116.pdf>.
- [4]. D.R. Cox, 1984 "*Analysis of Survival Data*", Chapman Hall, New York
- [5]. S. Galbriath, 2004, "Statistics for biomedical research", *Course Notes*, Australian Mathematical Institute, Summer School, University of New South Wales, Australia, 2004.
- [6]. Statsoft, 2006. "*Survival Analysis*", <http://www.statsoft.com/textbook/stsurvan.html>
- [7]. Wikimedia, 2006. *Survival Analysis*. http://en.wikimedia.org/wiki/survival_analysis
- [8]. D. Zhang, 2005. *Analysis of Survival Data*.
<http://www4.stat.ncsu.edu/~dzhang2/st745/chap4.pdf>