

**PENYELESAIAN MODEL SURVIVAL DENGAN HAZARD NON-  
PROPORSIONAL****Rahmat Hidayat<sup>1\*</sup>, Titik Pitriani Muslimin<sup>2</sup>, Marwan Sam<sup>3</sup>**<sup>1,3</sup> Program Studi Matematika, Fakultas Sains, Universitas Cokroaminoto Palopo<sup>2</sup>Program Studi Matematika, Fakultas Pendidikan Matematika, Universitas Cokroaminoto  
Palopo

\*E-mail: dayatmath@gmail.com

**ABSTRAK**

Analisis survival adalah suatu metode statistika yang digunakan untuk analisis data pada durasi daya tahan. Waktu survival adalah jangka waktu yang dicapai suatu objek dari waktu awal (waktu pemberian perlakuan) sampai terjadinya suatu peristiwa (*event*). Peristiwa ini dapat berupa lama waktu studi, pengembangan penyakit, respon terhadap pengobatan, kambuh, atau kematian. Waktu dari setiap individu tidak harus sama dan satuannya dapat berupa tahun, bulan, minggu atau hari. Durasi ini tidak semua dapat diamati secara lengkap. Hal ini bisa disebabkan oleh berakhirnya masa pengamatan, individu yang diamati menghilang dalam masa pengamatan, meninggal dunia atau sebab lain diluar penelitian. Apabila durasi ini dapat diamati selama masa pengamatan maka disebut data lengkap, sedangkan bila tidak dapat diamati secara lengkap maka disebut data tersensor. Salah satu model dalam analisis survival adalah Cox proporsional hazard. Model ini mengisyaratkan bahwa setiap kovariat bersifat proporsional. Kenyataannya, sering ditemukan kasus dimana terdapat kovariat yang tidak proporsional, dalam arti ada keterkaitan dengan peringkat waktu. Sehingga dalam hal ini perlu dilakukan perluasan model.

*Kata Kunci: Analisis survival, cox proporsional hazard, tersensor, tidak proporsional*

**Latar Belakang**

Dalam kehidupan sehari-hari terdapat banyak masalah yang responsnya berupa waktu ketahanan (*survival time*) atau sering juga disebut waktu kesintasan. Misalnya waktu yang diperlukan oleh pasien untuk sembuh dari penyakitnya, waktu sampai timbulnya reaksi atas suatu perlakuan, dan waktu yang diperlukan

oleh mahasiswa untuk memperoleh pekerjaan pertama setelah lulus. Waktu ketahanan juga dapat berupa suatu hal negatif yaitu waktu kegagalan (*failure*), misalnya durasi waktu rusaknya alat elektronik atau durasi waktu pasien penyakit tertentu dapat bertahan hidup. Lee (1992) menyatakan bahwa waktu survival didefinisikan sebagai waktu sampai

terjadinya suatu peristiwa. Data pengamatan waktu survival disebut data survival. Umumnya data survival tidak lengkap, artinya waktu survival tidak diketahui secara tepat karena terbatasnya waktu penelitian dan lain-lain. Hal inilah yang menyebabkan distribusi dari waktu survival menjadi tidak normal melainkan condong ke kanan (*positively skewed*) sehingga diperlukan suatu metode yang memfasilitasi ketidaknormalan data survival yaitu analisis survival (*survival analysis*).

Analisis survival merupakan metode untuk menganalisis data berupa waktu hingga mencapai suatu kejadian atau *event* tertentu. Pada proses pengumpulan data di lapangan, seringkali peneliti tidak memperoleh data yang lengkap dengan berbagai alasan. Maka muncullah istilah data tersensor untuk menangani hal tersebut. Data dikatakan tersensor jika hanya diketahui sebagian informasi mengenai waktu hingga terjadinya suatu *event* pada individu, namun tidak diketahui secara pasti waktu kejadiannya. Sebagian besar data di

lapangan merupakan data tersensor. Oleh karena itu dalam analisis survival dikembangkan berbagai metode untuk menganalisis data tersensor.

### **Analisis Survival**

Dalam perkembangan analisis data survival, dikenal istilah sensor atau penyensoran. Penyensoran dilakukan untuk mengatasi ketidaklengkapan suatu data pengamatan. Data dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena individu atau subjek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian subjek tersebut belum mengalami kejadian tertentu, sedangkan data yang dapat diamati secara lengkap sampai penelitian berakhir disebut data yang tidak tersensor (Lee & Wang, 2003). Menurut Kleinbaum Dan Klein (2005) tiga penyebab data dikatakan tersensor antara lain:

- a. *Loss to follow up*, yaitu subjek menghilang selama masa pengamatan, missal subjek pindah atau menolak untuk diamati.
- b. Subjek tidak mengalami kejadian selama penelitian.

- c. Subjek terpaksa diberhentikan dari pengamatan karena meninggal sebelum pengamatan berakhir atau alasan lain.

Menurut Collet (1994) dalam analisis survival terdapat tiga tipe penyensoran yaitu sensor kanan, sensor kiri serta sensor interval. Berikut adalah penjelasannya:

1. Sensor kanan (*right censoring*)

Sensor yang terjadi dikarenakan objek pengamatan belum mengalami kejadian hingga akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh. Misalkan suatu individu diamati selama lima tahun dari awal pengamatan, kemudian pada tahun ketiga individu tersebut pindah ke negara lain dan tidak dapat diamati lagi (*lost to follow up*). Individu ini memiliki waktu survival dalam penelitian setidaknya dua tahun, sehingga waktu pengamatan individu tersebut dikatakan tersensor kanan.

2. Sensor kiri (*left censoring*)

Sensor yang terjadi dikarenakan waktu awal dari subjek

pengamatan tidak dapat teramati pada awal pengamatan, sementara kegagalan dapat diamati secara penuh sebelum penelitian berakhir. Sebagai contoh, peneliti mengamati pasien penyakit kanker, peneliti dapat mencatat kejadian tepatnya seseorang tersebut positif kanker di tes pertamanya, namun peneliti tidak memiliki catatan tentang waktu tepatnya seseorang tersebut mulai berpenyakit kanker, dengan demikian pasien kanker tersebut tersensor kiri yaitu ketika mengalami kejadian pertama dengan hasil positif kanker.

3. Sensor interval (*interval censoring*)

Sensor interval adalah sensor yang waktu survivalnya berada dalam suatu selang tertentu. Sebagai contohnya, jika catatan medis menunjukkan bahwa pada usia 45 tahun pasien kanker dalam contoh di atas kondisinya sehat dan belum berpenyakit kanker, kemudian pasien melakukan tes pertama saat berumur 50 tahun dan terdiagnosis terkena penyakit kanker, dengan demikian usia

saat didiagnosis positif kanker adalah antara 45 dan 50 tahun.

Fungsi *survival* adalah peluang individu dapat bertahan hingga atau lebih dari waktu  $t$  yang didefinisikan sebagai berikut:

**Fungsi Survival**

$$\begin{aligned} S(t) &= P(T > t) \\ &= 1 - P(T \leq t) \\ &= 1 - F(t) \end{aligned} \tag{1}$$

Dimana  $F(t) = P(T \leq t)$ . Fungsi *survival* juga adalah integral fungsi kepekatan peluang  $f(t)$

$$\begin{aligned} S(t) &= P(T > t) \\ &= \int_t^{\infty} f(u) du \end{aligned} \tag{2}$$

**Fungsi Kepekatan Peluang**

peluang individu mengalami kejadian dalam interval  $t$  sampai  $\delta t$ .

Fungsi kepekatan peluang didefinisikan sebagai limit dari

$$f(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t)}{\delta t} \tag{3}$$

**Fungsi Hazard**

kegagalan atau meninggal pada waktu  $t$  dengan syarat bahwa seseorang itu telah bertahan hingga waktu  $t$ , fungsinya diberikan Cox pada tahun 1972:

Fungsi hazard yaitu fungsi yang menyatakan peluang seseorang mengalami risiko atau kejadian seperti

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \tag{4}$$

Dari definisi di atas diperoleh hubungan antara fungsi *survival* dengan fungsi *hazard*. Dengan menggunakan definisi peluang bersyarat, diperoleh:

$$\begin{aligned}
 h(t) &= \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \\
 &= \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t) / P(T \geq t)}{\delta t} \\
 &= \frac{\left( \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t)}{\delta t} \right)}{1 - F(t)} \\
 &= \frac{f(t)}{1 - F(t)} \\
 h(t) &= \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t)
 \end{aligned} \tag{5}$$

Persamaan (5) diintegrasikan dari 0 sampai  $t$  dengan  $S(0)=1$  yaitu

$$\begin{aligned}
 -\int_0^t h(x) dx &= \ln S(t) \\
 H(t) &= -\ln S(t) \\
 S(t) &= \exp[-H(t)]
 \end{aligned} \tag{6}$$

### Model Cox Proporsional Hazard

Cox dan Oakes (1984) menyatakan dalam analisis data survival, metode parametrik digunakan apabila bentuk distribusi survival-nya diketahui. Meskipun demikian dalam prakteknya bentuk pasti dari distribusinya kadang tidak diketahui dan kita mungkin tidak menemukan

model yang tepat. Oleh karena itu, menggunakan metode parametrik untuk melakukan analisis terhadap data survival sangat terbatas, sehingga dikembangkan model baru untuk menangani hal tersebut. Model Cox proporsional *hazard* adalah model yang biasa digunakan. Model ini tidak memerlukan pengetahuan tentang bentuk distribusi.

Cox proporsional *hazard* adalah model yang biasa digunakan untuk pendekatan multivariat untuk analisis data. Model ini memiliki ciri bahwa individu yang berbeda memiliki fungsi *hazard* yang proporsional yakni  $[h(t|x_1)/h(t|x_2)]$ , rasio fungsi *hazard* dari dua individu dengan penyertaan kovariat  $x_1 = (x_{11}, x_{21}, \dots, x_{p1})'$  dan

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (7)$$

dengan  $x$  adalah kovariat, tetapi ia tidak membuat asumsi tentang bentuk dari  $h_0(t)$  yang disebut dengan

$$\theta = \exp\{\beta_0 + \beta_1 x\} \quad (8)$$

$\beta_0$  dimasukkan ke dalam baseline fungsi *hazard*  $h_0$ . Fungsi *hazard* untuk beberapa pasangan kovariat yang

$$HR = \frac{h_0 \exp\{\beta x_i\}}{h_0(t) \exp\{\beta x_j\}} = \exp\{\beta(x_i - x_j)\} \text{ untuk } i \neq j$$

Jika responden mempunyai beberapa karakteristik, metode Cox proporsional *hazard* dapat

$x_2 = (x_{12}, x_{22}, \dots, x_{p2})'$  adalah konstan. Ini artinya bahwa rasio dari resiko kegagalan dari dua individu adalah sama tidak bergantung pada seberapa lama mereka bertahan. Cox (1972) menjelaskan bahwa bentuk umum dari model Cox proporsional *hazard* adalah:

*baseline* fungsi *hazard* karena itu adalah nilai dari fungsi *hazard* saat  $x = 0$ . Ketika menggunakan kovariat

berbeda  $i$  dan  $j$  dapat dibandingkan dengan menggunakan *hazard* rasio.

menerangkan pengaruh karakteristik-karakteristik tersebut terhadap peubah respon secara simultan.

Tabel 1. Penduga parameter, nilai- $p$ , dan *hazard ratio* dengan menggunakan model Cox proporsional *hazard*

Peubah penjelas	Penduga parameter	Nilai - $p$	<i>Hazard ratio</i>
Jenis Kelamin	-0.32850	0.3871	0.720
Status Menikah (1)	0.53532	0.0077*	1.708
Umur	0.97377	0.7450	2.648
Status Kerja	1.47270	<0.0001*	4.361
Ekonomi Sedang (1)	0.83471	<0.0001*	0.434
Ekonomi Atas (2)	-0.52923	<0.0001*	0.589
Jumlah Saudara	0.15443	0.6542	1.167

Pada Tabel 2 terlihat bahwa peubah penjelas yang berpengaruh nyata terhadap lama studi mahasiswa adalah status menikah, status bekerja, status ekonomi sedang dan atas.

Selanjutnya dilakukan penaksiran terhadap asumsi

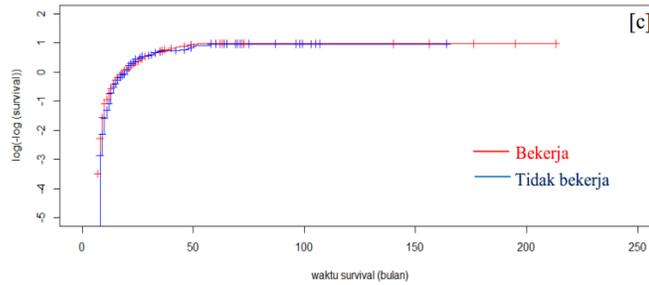
proporsional untuk setiap kovariat. Dalam tulisan ini hanya digunakan metode Schoenfeld residuals untuk memeriksa kovariat yang tidak memenuhi asumsi proporsional pada lama studi mahasiswa. Hasilnya disajikan sebagai berikut

**Tabel 2** Korelasi dan nilai  $-p$  peubah penjelas

Peubah penjelas	Korelasi	Nilai $-p$
Jenis Kelamin	0.05438	0.135
Status Menikah (1)	0.02557	0.231
Umur	-0.05312	0.543
Status Kerja	0.33078	0.001*
Ekonomi Sedang (1)	0.01169	0.713
Ekonomi Atas (2)	-0.10459	0.289
Jumlah Saudara	-0.04565	0.225

Nilai  $-p$  kovariat umur kurang dari 0.05 yang berarti terdapat korelasi antara kovariat tersebut dengan peringkat waktu sampai mahasiswa menyelesaikan kuliah sehingga kovariat status kerja tidak memenuhi

asumsi proporsional. Hal ini juga bisa dilihat dari grafik fungsi *log minus log* (LML). Grafik/plot yang berpotongan atau berhimpit menunjukkan adanya interaksi antara peubah bebas terhadap waktu *survival* (tidak proporsional).



**Gambar 1** Grafik fungsi LML untuk peubah status bekerja

Sehingga data kembali dimodelkan dengan model *Cox extended*. Sebelum dilakukan analisis, terlebih dahulu dilakukan uji untuk melihat bentuk  $g_j(t)$  yang lebih sesuai. Uji ini

menggunakan kriteria AIC. Menurut metode AIC, model regresi terbaik adalah model yang mempunyai nilai AIC terkecil (Love, *et al.* 2003).

**Tabel 3.** Nilai AIC model

No.	Jenis Model	Nilai AIC
1.	Cox Proportional Hazard <i>Cox extended</i>	377.639
2.	$g_1(t) = t$	373.014
3.	$g_1(t) = \log t$	371.802

Nilai AIC dengan bentuk  $g_1(t) = \log t$  adalah yang paling kecil diantara dua model lainnya. Sehingga model inilah yang terbaik.

Selanjutnya dilakukan uji untuk melihat kovariat yang berpengaruh nyata terhadap peubah respon. Hasilnya disajikan sebagai berikut

**Tabel 4.** Penduga parameter, nilai- $p$ , dan *hazard ratio* dengan menggunakan model *Cox extended*

Peubah penjelas	Penduga parameter	Nilai - $p$	<i>Hazard ratio</i>
Jenis Kelamin	-0.4780	0.3761	0.620
Status Menikah (1)	0.4774	0.0077*	1.612
Umur	0.9470	0.7450	2.578
Status Kerja	1.3668	<0.0001*	3.923
Ekonomi Sedang (1)	-0.8843	<0.0001*	0.413
Ekonomi Atas (2)	-0.7657	<0.0001*	0.465
Jumlah Saudara	0.1544	0.6542	1.167
Status Kerja.lt	0.6392	<0.0001*	1.895

Pada Tabel 5 terlihat bahwa peubah penjelas yang berpengaruh nyata lama studi mahasiswa adalah status menikah, status bekerja, status ekonomi sedang dan atas.

Model Cox extended yang melibatkan variabel yang signifikan mempengaruhi lama studi mahasiswa adalah sebagai berikut:

$$\hat{h}(t, \mathbf{x}(t)) = \hat{h}_0(t) \exp(0.4774 \text{ Status Menikah} + 1.3668 \text{ Status Kerja} - 0.8843 \text{ Ekonomi sedang (1)} - 0.7657 \text{ Ekonomi atas (2)} + 0.6392 \text{ Status kerja} \times \log(t))$$

### **Interpretasi**

Tabel 5 menunjukkan bahwa peubah status menikah, status bekerja, status ekonomi sedang dan atas merupakan peubah yang berpengaruh nyata terhadap lama studi mahasiswa. Dari nilai rasio *hazard* terlihat bahwa mahasiswa yang belum menikah memiliki peluang untuk lulus lebih cepat dibandingkan mahasiswa yang telah menikah sebesar 1.612 kali. Hal ini diakibatkan karena setelah menikah, maka tanggung jawab mahasiswa akan semakin banyak. Hal tersebut membuat konsentrasi mahasiswa terbagi antara fokus untuk kuliah dan tanggung jawab terhadap keluarga. Begitupun mahasiswa yang belum bekerja memungkinkan untuk

menyelesaikan kuliah lebih cepat dibandingkan dengan mahasiswa yang bekerja sebesar 1.895 kali. Mahasiswa yang bekerja cenderung bermasalah dalam kuliah, diantaranya adalah masalah kehadiran saat kuliah, hal ini tentu saja berpengaruh terhadap tugas kuliah dan sebagainya, sehingga nilai mahasiswa tersebut juga cenderung kurang. Adanya nilai yang kurang ini membuat mahasiswa wajib untuk mengulang matakuliahnya sehingga memperpanjang masa studinya. Status ekonomi orang tua mahasiswa dengan status rendah lebih lambat menyelesaikan kuliah dibandingkan dengan mahasiswa yang orang tuanya memiliki status ekonomi sedang dan rendah berturut-turut sebesar 0.413 dan 0.465, hal ini berdampak langsung

terhadap ketersediaan sarana dan prasarana penunjang kuliah.

#### DAFTAR PUSTAKA

- Allison PD. 2010. *Survival Analysis Using SAS: A Practical Guide*. 2<sup>nd</sup>Ed. USA: SAS Institut INC.
- Barker K dan Baroud H. 2014. Proportional hazards models of infrastructur system recovery. *Reliability Engineering and System Safety*. 124:201-206.
- Bradburn M, Clark TG, Love C. 2003. Multivariate data analysis, an introduction to concepts and methods. *British Journal of Cancer*. 89(3): 431-436.
- Collet D. 2003. *Modelling Survival Data in Medical Research* 2<sup>nd</sup>Ed. London: Chapman & Hall/CRC.
- Cox DR. 1972. Regression models and life tables (with discussion). *J R Statisc Sic B*. 34(2): 187-220.
- Devarajan K dan Ebrahim N. 2010. A semi-parametric generalization of the Cox proportional hazards regression model: Inference and applications. *Computational Statistics and Data Analysis*. 55:667-676.
- Guillory, Christopher W. 2008. A Multilevel Discrete- Time Hazard Model Of Retention Data In Higher Education, Dissertation, Louisiana State University, Louisiana.
- Hidayat R, Sumarno H, Nugrahani EH. 2014. *Survival Analysis in Modeling the Birth Interval of the First Child in Indonesia*. *Open Journal of Statistics*. 4:198-206
- Hogg VR, Craig TA. 1995. *Introduction to Mathematical Statistics*. 5<sup>nd</sup>Ed. New Jersey: Prentice Hall, Englewood Cliffs Publisher.
- Kaplan EL, Meier P. 1958. Non Proportional Estimation from Incomplete Observation. *Journal of the America Statistical Association*. 53: 457-481
- Lawless JF. 2002. *Statistical Models and Methods for Lifetime Data*. 2<sup>nd</sup>Ed. Canada: Wiley Series In Probability and Statistics.

- Lee ET, Wang JW. 2003. *Statistical Methods for Survival data Analysis. 3<sup>rd</sup>Ed.* New York: A Wiley Interscience Publication.
- Li, Y dan Ryan L. 2002. Modeling Spatial Survival Data Using Semiparametrik Frailty Models. *Biometrics*. 58: 287-297.
- Love C, Altman DG, Bradburn M. 2003. Multivariate data analysis. *British Journal of Cancer*, 89(3): 437-443.
- Rahayu, P. 2003. *Regresi Survival Hazard Proporsional Cox Sebagai Metode Alternatif Bagi Regresi Logistik Biner Dalam Mengidentifikasi Faktor Resiko. Studi Kasus Kematian Penderita Penyakit Jantung Koroner di RSUD Dr. Soetomo Surabaya*, Thesis, Surabaya.
- Silva GO, Ortega EM, Paula GA. 2011. Residuals for log-Burr XII regression models in survival analysis. *Journal of Applied Statistics*. 38(7):1435-1445
- Zhang Y. 2008. *Parametrik Mixture Models in Survival Analysis with Applications*. Temple University: Philadelphia.