

Analisis Survival Dengan Model Regresi Cox

Study Kasus: Pasien Demam Berdarah Dengue di Rumah Sakit Haji Surabaya

Ni Putu Lisa Ernawatiningsih

Institut Teknologi Sepuluh November, Sukolilo, 16680 Surabaya, Indonesia

e-mail: ernawatiningsih.lisa@gmail.com

Purhadi

Institut Teknologi Sepuluh November, Sukolilo, 16680 Surabaya, Indonesia

e-mail: purhadi@statistika.its.ac.id

Abstract: Demam berdarah dengue (DBD) merupakan suatu wabah penyakit musiman yang jika tidak mendapatkan penanganan yang tepat akan bisa menjadi sebuah kejadian luar biasa, karena penyebarannya sangat cepat dan dapat menyebabkan kematian. Untuk mengurangi angka kematian akibat demam berdarah dengue, maka penelitian ini akan memodelkan waktu survival dengan studi kasus pada pasien demam berdarah dengue yang dirawat di Rumah Sakit Haji Surabaya pada tahun 2011. Metode yang digunakan adalah Analisis *Survival* dengan model regresi cox. Adapun hasil analisisnya berdasarkan nilai AIC masing-masing faktor, diperoleh faktor yang paling berpengaruh terhadap laju kesembuhan pasien adalah usia dan jumlah trombosit. Dari 66 pasien yang dirawat 50% pasien berusia 0-14 tahun sisanya diatas 14 tahun, dan 70% pasien yang jumlah trombosit di bawah normal sisanya normal. Dari nilai odds ratio untuk usia dan jumlah trombosit disimpulkan pasien yang berusia lebih tua akan lebih lama kesembuhannya dari pada pasien yang berusia lebih muda dan ketika jumlah trombosit di bawah normal akan mencapai sembuh lebih lama daripada pasien dengan jumlah trombosit normal. **Keywords:** Demam Berdarah Dengue, Regresi Cox, Analisis *Survival*, Distribusi Lognormal

1. Pendahuluan

Analisis statistik yang mencari hubungan antara variabel respon dengan variabel prediktor salah satunya adalah dengan menggunakan metode regresi. Jika variabel respon yang digunakan adalah berupa data survival (waktu), analisis yang digunakan adalah analisis survival [1]. Salah satu analisis *survival* digunakan regresi Cox. Regresi Cox pertama kali dikembangkan oleh Cox pada tahun 1972. Regresi ini, lebih populer digunakan dalam penelitian tentang data kesehatan, data ekonomi yang variabel responnya berupa waktu (hari, bulan, tahun). Misalnya data tentang waktu pasien menderita penyakit tertentu, dimana dimulai dari awal masuk rumah sakit sampai terjadi kejadian tertentu, seperti kematian, sembuh atau kejadian khusus lainnya [9]. Regresi cox ini tidak mempunyai asumsi mengenai sifat dan bentuk yang sesuai dengan distribusi normal seperti asumsi pada regresi yang lain, distribusi yang digunakan adalah sesuai dengan distribusi dari variabel responnya, yang diperoleh dari uji Anderson-Darling [2]. Salah satu masalah utama dalam analisis survival data untuk mempelajari ketergantungan waktu kelangsungan hidup dalam variabel prediktor. Hal ini sering disederhanakan

dengan menggunakan model Cox proportional hazard. Tujuan dari penelitian ini adalah untuk memperkirakan parameter dan pengujian hipotesis untuk regresi Cox Lognormal dua parameter.

2. Tinjauan Pustaka

2.1. Analisis *Survival*

Analisis *survival* merupakan suatu metode statistik yang berkaitan dengan waktu, yaitu dimulai dari *time origin* atau *start point* sampai pada suatu kejadian khusus (*failure event/end point*). Salah satu analisis *survival* yang digunakan adalah regresi cox, yaitu suatu regresi yang digunakan untuk analisis data dengan variabel dependennya berupa waktu *survival*. Waktu *survival* yang dicatat adalah satu dari dua kriteria berikut:

Selisih waktu mulai dilakukannya pengamatan sampai waktu terjadinya kematian atau disebut data tidak tersensor (*uncensored data*). Jika waktu kematiannya tidak diketahui, maka dipakai selisih waktu dari mulai dilakukannya pengamatan sampai waktu akhir penelitian disebut data tersensor (*censored data*).

Beberapa kemungkinan penyebab kematian data tersensor adalah masa penelitian berakhir sementara observasi masih belum mencapai *failure*. Kematian karena sebab tertentu disebut *failure event* [3]. Suatu pengamatan waktu survival tidak sampai pada *failure event*, penyebabnya adalah *Lost of follow up*, *drop out*, dan *termination of the study*.

Fungsi *survival* digunakan untuk menyatakan probabilitas suatu individu bertahan dari waktu mula-mula sampai waktu t . Waktu survival dilambangkan dengan T yang merupakan variabel random dan mempunyai fungsi distribusi peluang $f(t)$. Fungsi *survival* $S(t)$, didefinisikan sebagai probabilitas bahwa waktu *survival* lebih besar atau sama dengan t sehingga:

$$S(t : \mu, \sigma) = 1 - \Phi \left(\frac{\ln(t) - \mu}{\sigma} \right) \quad (1)$$

Fungsi hazard menyatakan proporsi atau laju kematian seketika suatu individu yang survive sampai waktu ke- t . Berikut adalah fungsi hazard tanpa pengaruh variabel bebas yang biasa disebut dengan fungsi *baseline hazard*:

$$h(t : \mu, \sigma) = \frac{\Phi \left(\frac{\ln(t) - \mu}{\sigma} \right)}{\left[1 - \Phi \left(\frac{\ln(t) - \mu}{\sigma} \right) \right] \sigma(t)} \quad (2)$$

2.2. Regresi Cox Lognormal

Fungsi densitas peluang dari distribusi lognormal dua parameter adalah sebagai berikut [4]:

$$f(t, \mu, \sigma) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi t}} \exp\left\{-\frac{[\log(t) - \mu]^2}{2\sigma^2}\right\}, & \sigma > 0, \mu > 0 \\ 0, & \text{untuk yang lainnya} \end{cases} \quad (3)$$

dimana μ adalah parameter skala (*scale*) dan σ adalah parameter bentuk (*shape*).

Regresi cox merupakan salah satu analisis survival yang sering digunakan, metode ini pertama kali dikenalkan oleh Cox dan respon yang digunakan adalah data yang diperoleh dari perhitungan waktu suatu peristiwa tertentu (waktu survival) [5]. Misalnya data tentang waktu pasien menderita penyakit tertentu, dimana perhitungannya dimulai dari awal sakit hingga terjadi pada kejadian khusus, yaitu seperti kematian, sembuh atau kejadian khusus lainnya [9]. Regresi cox ini tidak mempunyai asumsi mengenai sifat dan bentuk sesuai dengan distribusi normal seperti asumsi pada regresi yang lain, distribusi yang digunakan adalah sesuai dengan respon yang digunakan [6]. Model dari regresi cox adalah sebagai berikut:

$$h(t) = h_0(t) \exp(\beta'x) \quad (4)$$

dimana $h_0(t)$ merupakan fungsi *baseline hazard* yang diperoleh dari fungsi *hazard* pada distribusi Lognormal dua parameter. Bentuk umum fungsi baseline hazard dari distribusi Lognormal dua parameter adalah sebagai berikut [7]:

$$\hat{h}(t_i) = \frac{\Phi\left(\frac{\ln(t_i) - \hat{\mu}}{\hat{\sigma}}\right)}{\left[1 - \Phi\left(\frac{\ln(t_i) - \hat{\mu}}{\hat{\sigma}}\right)\right] \hat{\sigma}(t_i)} \exp(\beta'X_i) \quad (5)$$

2.3. Estimasi Parameter dan Statistik Uji

Menentukan estimasi parameter untuk fungsi baseline hazard dilakukan dengan metode Maximum Likelihood Estimator (MLE). Dengan fungsi likelihood sebagai berikut:

$$L(t; \mu, \sigma) = \prod_{i=1}^n \left[\frac{1}{\sigma\sqrt{2\pi(t)}} \exp\left\{-\frac{[\log(t) - \mu]^2}{2\sigma^2}\right\} \right] \quad (6)$$

untuk mempermudah mendapatkan nilai penaksir parameter dapat digunakan metode iterasi Newton-Raphson. Sedangkan menentukan model terbaik dibutuhkan estimasi terhadap koefisien variabel bebas dengan menggunakan model Maximum Partial Likelihood Estimator (MPLE). Maximum Partial Likelihood Estimator (MPLE) digunakan untuk menghilangkan parameter-parameter pengganggu selain parameter regresi. Dengan fungsi likelihoodnya sebagai berikut:

$$L(B) = \left[\prod_{i=1}^r \frac{\exp(\beta^T z_{(i)})}{\sum_{l \in \mathfrak{R}(t_{(i)})} \exp(\beta^T z(l))} \right]^{1-\delta_i} \quad (7)$$

dimana $\mathfrak{R}(t_{(i)})$ adalah himpunan resiko pada waktu t_i dan δ_i bernilai 0 apabila data tidak tersensor dan 1 untuk data tersensor, untuk mempermudah mendapatkan nilai penaksir parameter dapat digunakan metode iterasi Newton-Raphson [8].

Uji parameter dilakukan untuk mengetahui signifikansi parameter terhadap respon atau model. Uji parameter ada dua yaitu:

- (a) Uji serentak ini digunakan untuk mengetahui apakah variabel independen yang digunakan pada model berpengaruh signifikan secara bersama-sama.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0, \text{ dengan } k = 1, 2, \dots, p$$

$$G^2 = -2 \ln \frac{L(\hat{\omega})}{L(\hat{\Omega})} \quad (8)$$

Keputusan : tolak H_0 jika nilai statistik uji $G^2 > \chi_{p,\alpha}^2$

- (b) Uji Parsial untuk mengetahui variabel independen mana yang berpengaruh signifikan terhadap model, maka langkah selanjutnya adalah melakukan uji parsial terhadap parameter-parameternya. Berikut hipotesisnya:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

$$W_k = \left[\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right]^2 \quad (9)$$

Keterangan:

$\hat{\beta}_k$ adalah penduga β_k

$SE(\hat{\beta}_k)$ adalah standar error dari β_k

Keputusan : Tolak H_0 jika $\chi_{hitung}^2 > \chi_{l,\alpha/2}$

2.4. Demam Berdarah Dengue (DBD)

Penyakit Demam Berdarah Dengue (DBD) adalah penyakit menular yang disebabkan oleh virus dengue dan ditularkan oleh nyamuk *Aedes aegypti*. Beberapa tes yang biasanya dilakukan adalah studi serologi untuk mengetahui ada tidaknya antibodi terhadap virus dengue di tubuh pasien, menghitung jumlah antibodi terhadap virus dengue, dan penghitungan darah lengkap (hemoglobin, leukosit, hematokrit, dan trombosit) [9].

Hemoglobin (Hb) merupakan suatu protein kompleks yang mengandung zat besi dan berwarna merah yang ada dalam eritrosit. Hemoglobin ini sangat diperlukan oleh tubuh, apabila kadar hemoglobin kurang, maka kondisi seseorang akan lemah. Kadar hemoglobin yang normal untuk wanita adalah 11,4 sampai 15,1 g/dl, sedangkan untuk laki-laki adalah 13,4 sampai 17,7 g/dl. Leukosit atau sel darah putih adalah sel darah yang mengandung inti, dengan batas normal bagi wanita adalah 4300-11300/mm³, untuk laki-laki adalah 4300-10300/mm³. Hematokrit adalah suatu angka yang menunjukkan prosentase zat padat dalam darah terhadap cairan darah. Keadaan normal hematokrit pada tubuh manusia adalah antara 38 sampai 42% untuk wanita dan 40-47% untuk laki-laki. Trombosit, merupakan bagian terkecil dari unsur seluler pada sumsum tulang dan penting dalam proses pembekuan serta hemostatis. Dalam kondisi normal, jumlah trombosit untuk wanita dan laki-laki adalah antara 150000-400000/mm³.

3. Aplikasi Analisis Survival pada Kasus Demam Berdarah Dengue

3.1. Sumber Data

Sumber data pada penelitian ini adalah data rekam medis dari Rumah Sakit Haji Surabaya mengenai waktu survival pasien demam berdarah dengue pada tahun 2011. Waktu survival adalah lamanya pasien demam berdarah dirawat dari awal sampai dinyatakan sembuh oleh pihak Rumah Sakit Haji Surabaya. Unit observasi yang digunakan dalam penelitian ini sebanyak 66 pasien penderita demam berdarah dengue di Rumah Sakit Haji Surabaya. Pasien yang keluar dari rumah sakit tanpa izin dari pihak rumah sakit tidak digunakan dalam penelitian ini.

Secara ringkas variabel-variabel yang digunakan dalam penelitian ini dapat dilihat pada tabel 1 berikut:

Tabel 1. Variabel Penelitian

Variable	Remark	Type	Category
T	Lamanya (waktu survival) pasien DBD dirawat di RS Haji, mulai dirawat sampai dinyatakan sembuh	Kontinu	–
X_1	Usia pasien DBD di RS Haji dirawat awal masuk	Kontinu	–
X_2	Jenis kelamin pasien DBD di RS Haji	Kategorik	1 : laki-laki 2 : perempuan
X_3	Jumlah hemoglobin pasien DBD saat diperiksa pertama kali masuk di RS	Kontinu	–
X_4	Jumlah leukosit pasien DBD saat diperiksa pertama kali masuk di RS	Kontinu	–
X_5	Persentase hematokrit pasien DBD saat diperiksa pertama kali masuk di RS Haji	Kontinu	–
X_6	Jumlah Trombosit pasien DBD saat diperiksa pertama kali masuk di RS	Kontinu	1 : dibawah normal 2 : normal

3.2. Uji Distribusi Data

Pengujian distribusi data dapat digunakan untuk mengetahui distribusi yang tepat untuk variabel prediktor [8]. Uji yang digunakan pada penelitian ini adalah uji Anderson-Darling, berikut adalah hasilnya:

Tabel 2. Pengujian Distribusi Data

Distribusi	Anderson Darling (adj)
Lognormal (2P)	1.889
Smallest extreme value	2.040
Exponential (2P)	9.062
Exponential	15.395

Tabel 2 menunjukkan bahwa distribusi lognormal memiliki nilai terkecil, sehingga dapat dikatakan bahwa variabel respon yang berupa *survival time* mendekati distribusi Lognormal.

3.3. Faktor-faktor yang Berpengaruh Terhadap Laju Kesembuhan pasien DBD

Tabel 3 dibawah ini menunjukkan variabel bebas yang secara signifikan berpengaruh terhadap laju kesembuhan pasien DBD ($\alpha = 0.05$). Berikut adalah hasilnya:

Table 3. Uji Wald Chi-Square untuk variabel prediktor

Variabel	Db	Wald Chi-Square	Sig.
Umur	1	5.6929	0.0170
Jenis Kelamin	1	0.4123	0.5208
Hemoglobin	1	2.7414	0.0978
Leukosit	1	0.0226	0.8805
Hematokrit	1	3.5004	0.0614
Trombosit	1	5.8987	0.0152

Uji serentak dilakukan untuk mengetahui pengaruh variabel prediktor secara bersama-sama, berikut hipotesisnya:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{at least one } \beta_k \neq 0, \text{ with } k = 1, 2, \dots, 6$$

$$\alpha : 0.05$$

$$G^2 = 17.248$$

$$\chi_{6,\alpha}^2 = 1.635$$

Keputusan: Tolak H_0 , karena nilai $G^2 > \chi_{6,\alpha}^2$. Jadi paling tidak ada satu variabel prediktor yang berpengaruh signifikan terhadap model. Uji Parsial dilakukan untuk mengetahui pengaruh masing-masing variabel prediktor, berikut hipotesisnya:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

$$\alpha : 0.05$$

$$\chi_{1,\alpha}^2 = 0.003932$$

Keputusan: Tolak H_0 karena nilai $\chi_{hitung} > \chi_{1,\alpha}^2$. Hasilnya dapat dilihat pada Tabel 3.

Tabel 4. Nilai AIC untuk Model Terbaik

Variable	AIC
Semua variabel prediktor	12.0004
Tanpa Leukosit	10.0004
Tanpa Leukosit dan Jenis Kelamin	8.0003
Tanpa Leukosit, Jenis Kelamin dan Hemoglobin	6.0001
Umur dan Trombosit	4.0001

Tabel 4. menunjukkan bahwa model terbaik adalah model dengan dua variabel prediktor, yaitu umur dan trombosit dengan nilai 4.0001.

Tabel 5. Cox regression parameter estimates on the best model

Variable	DF	Estimate	Wald Chi-Square	Sig.
Intercept	1	1.5875	575.85	< 0.0001
Umur	1	0.0055	4.4667	0.0346
Trombosit(1)	1	0.1541	4.9333	0.0263
Trombosit(2)	0	0.0000		
Parameter Scala	1	0.3278		
Parameter Bentuk	1	1.6554		

Dari hasil estimasi parameter pada Tabel. 5, model regresi Cox terbaik dapat dituliskan sebagai berikut

$$\hat{h}(t_i) = \frac{\Phi\left(\frac{\ln(t_i) - \hat{\mu}}{\hat{\sigma}}\right)}{\left[1 - \Phi\left(\frac{\ln(t_i) - \hat{\mu}}{\hat{\sigma}}\right)\right]} \exp(\beta' x_i)$$

$$\hat{h}(t_i) = \frac{\Phi\left(\frac{\ln(t_i) - 0.3278}{1.6554}\right)}{\left[1 - \Phi\left(\frac{\ln(t_i) - 0.3278}{1.6554}\right)\right]} \exp(0.0055 \text{ age} + 0.1541 \text{ platelets}(1))$$

Nilai estimasi parameter skala dan bentuk untuk masing-masing pasien adalah sama. Jika variabel umur meningkat satu (tahun) itu akan menaikkan fungsi *hazard* sebanyak $e^{0.0055}$. Begitupun dengan kenaikan trombosit sebesar 1 satuan akan meningkatkan fungsi *hazard* sebesar $e^{0.1541}$.

3.4. Laju Kesembuhan Pasien DBD

Untuk mengetahui laju kesembuhan pasien dapat dicari dengan menari nilai *odds ratio* variabel-variabel yang signifikan [8]. Berikut hasil *hazard ratio*:

Tabel 6. Nilai hazard ratio untuk umur dan trombosit dibawah normal

Variabel	Db	Estimasi (β)	<i>Hazard ratio</i> ($e^{\hat{\beta}}$)
Umur	1	0.0055	1.0055
Trombosit (1)	1	0.1541	1.1667

Tabel 6. menunjukkan risiko untuk sembuh pasien yang berusia satu tahun lebih tua adalah sebesar 1/1,0055 atau 0,9945 kali dari pasien yang lebih muda satu tahun. Dan laju kesembuhan pasien dengan jumlah trombosit di bawah normal dibandingkan dengan yang normal adalah 1/1,1667 atau 0,8572, maka laju kesembuhan pasien dengan jumlah trombosit di bawah normal adalah 0,8572 kali jumlah trombosit normal.

4. Kesimpulan

Dari nilai AIC terkecil yang diperoleh dapat disimpulkan bahwa faktor yang berpengaruh secara signifikan terhadap laju kesembuhan pasien DBD adalah umur dan trombosit. Dengan Model cox regresi (*hazard proportional model*) untuk masing-masing pasien adalah sebagai berikut:

$$\hat{h}(t_i) = \frac{\Phi\left(\frac{\ln(t_i)-0.3278}{1.6554}\right)}{\left[1 - \Phi\left(\frac{\ln(t_i)-0.3278}{1.6554}\right)\right]} \exp(0.0055 \text{ age} + 0.1541 \text{ platelets}(1))$$

Dari nilai *hazard ratio*, resiko pasien dengan umur yang lebih tua satu tahun cenderung memiliki laju kesembuhan yang lebih lama dari pasien yang lebih muda. Begitu juga untuk pasien yang memiliki jumlah trombosit dibawah normal (1) cenderung memiliki laju kesembuhan yang lebih lama dari pasien yang memiliki jumlah trombosit normal (2).

Daftar Pustaka

- [1] Liang, C.,Zheng, G., Zhu, N., Zhe, T., Lu, S., dan Chen, L.: A New Environmental Heat Stress Index for Indoor Hot and Humid Environments Based on Cox Regression. *Journal International of Buliding and Environment*, 46, 2472-2479 (2011)
- [2] Law, A. M., & Kelton, D. W.: *Simulation Modelling Analysis* (3th ed.). New York: MacGraw-Hill. (2000)
- [3] Le, C. T.: *Applied Survival Analysis*. New York: John Wiley and Sons, Inc. (1997)
- [4] Hill, B.M.: The three parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association* 58, 72-84 (1963)
- [5] Cox, D. R.: Regression Model and Life Table. *J Roy Stat Soc B*, 34, 187-202 (1972)
- [6] Ahmed, F. E., Vos, P. W., dan Holbert, D.: Modeling Survival in Colon Cancer : A Metodological Review. *Molecular Cancer*, 6, 15 (2007)
- [7] Hosmer, D.W., Lemeshow, S. and May, S.: *Applied Survival Analysis*. Wiley & Sons, Inc., Hoboken, New Jersey. (2008)
- [8] Collet, D.: *Modelling Survival Data in Medical Reseach*. London: Chapman and Hall. (1994)
- [9] Chuansumrit A, Tagnararatckakit K, Hongeng S, et al.: Dengue Hemorrhagic fever in children with underlying hematologic-oncologic diseases. *Thai J Hematol Transf Med*, 13, (2003)