



## PEMODELAN MULTILEVEL *SURVIVAL* DENGAN BAYESIAN *MARKOV CHAIN MONTE CARLO* PADA PENYAKIT DBD

Amaliyatul Mahfiyah✉, Arief Agoestanto

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia

Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang 50229

### Info Artikel

Sejarah Artikel:  
Diterima Juli 2021  
Disetujui November 2021  
Dipublikasikan November 2021

#### Keywords:

*Survival analysis, Multilevel model, Bayesian Markov Chain Monte Carlo, Demam Berdarah Dengue (DBD)*

### Abstrak

Penyakit DBD dapat mengakibatkan kematian dalam waktu yang singkat. Sampai saat ini penyakit DBD masih menjadi masalah kesehatan di Indonesia yang semakin meningkat jumlah penderitanya dan semakin luas penyebarannya. Tujuan dari penelitian ini adalah melakukan pemodelan multilevel *survival* pada data pasien Demam Berdarah *Dengue* (DBD) dan untuk mengetahui faktor-faktor yang mempengaruhi waktu kelangsungan hidup pasien DBD. Metode estimasi parameter yang digunakan dalam penelitian ini adalah metode *Bayesian Markov Chain Monte Carlo* dengan algoritma *Gibbs Sampling*. Untuk mempermudah dalam melakukan analisis data, peneliti menggunakan software WinBUGS. Hasil penelitian dari 90 sampel penderita demam berdarah *Dengue* di RSUD Tugurejo menunjukkan bahwa faktor yang berpengaruh secara signifikan terhadap waktu kelangsungan penderita DBD adalah variabel curah hujan, jumlah pasien perbulan, jenis kelamin, kejadian pendarahan dan jumlah trombosit. Pada hasil penelitian terdapat faktor yang tidak mempengaruhi waktu kelangsungan hidup penderita DBD. Maka pada penelitian selanjutnya diharapkan untuk bisa mengembangkan metode analisis *survival* lain seperti, *non proportional hazard*, model *failure*, model *mixture*, model *frailty* dan model lain.

### Abstract

*Dengue fever can cause death in a short time. Until now, DHF is still a health problem in Indonesia, where the number of sufferers is increasing, and its spread is wider. The purpose of this study was to model multilevel survival on patient data from Dengue Hemorrhagic Fever (DHF) and to determine the factors that affect the survival time of DHF patients. The parameter estimation method used in this study is the Bayesian Markov Chain Monte Carlo method with the Gibbs Sampling algorithm. To make it easier to perform data analysis, researchers used WinBUGS software. The results of the research from 90 samples of Dengue hemorrhagic fever patients at Tugurejo Hospital showed that the factors that significantly influenced the survival time of DBD sufferers were the variable rainfall, number of patients every month, gender, bleeding incidence and the number of trombosit. In the results of the study, there were factors that did not affect the lifetime of DHF patients. So, in future research, it is expected to develop other survival analysis methods such as non proportional hazard, failure model, mixture model, frailty model and other models.*

### How to cite:

Mahfiyah, A & Agoestanto, A. 2021. Pemodelan Multilevel *Survival* dengan Bayesian *Markov Chain Monte Carlo* pada Penyakit DBD. *UNNES Journal of Mathematics*. 10(2):1-11.

## PENDAHULUAN

Ilmu statistika berperan sangat penting dalam perkembangan ilmu pengetahuan dan teknologi. Hampir seluruh kebijakan yang diambil para pakar ilmu pengetahuan didasari oleh ilmu statistika. Dalam bidang kesehatan ilmu statistika yang sering digunakan adalah analisis *survival*. Analisis *survival* adalah suatu metode yang berhubungan dengan waktu, mulai dari awal pengamatan sampai suatu peristiwa terjadi (Lee & Wang, 2003). Pada analisis *survival* memerlukan data waktu ketahanan dari suatu individu.

Analisis regresi multilevel adalah metode yang dapat digunakan untuk mengatasi masalah dengan data berstruktur hirarki (Luh & Ayu, 2016). Pada model multilevel, struktur hirarki didefinisikan sebagai level. Menurut Harlan (2016) model multilevel dibedakan menjadi dua, yaitu model multilevel dengan *Nested Groupings* (pengelompokan tersarang) dan model multilevel dengan *non-Nested groupings* (pengelompokan tak tersarang). Sesuai persepektifnya model multilevel banyak direkomendasikan untuk menganalisis berbagai data berstruktur hirarki, seperti bidang psikologi (Silva, 2007), kesehatan (Yau, 2001; Ha & Lee, 2005; Raman & Hadeker, 2005), asuransi (Antonio *et al*, 2010), demografi (Lukoceine & Vermunt, 2008), kontrasepsi (Khan & Shaw, 2011). Menurut Hox sebagaimana dikutip oleh Mustafafi (2015), pemodelan multilevel berkaitan dengan populasi pada struktur hirarki atau bertingkat, dimana sampel dari populasi dapat digambarkan sebagai sampel multi-tahap (*multistage sample*).

Umumnya model regresi mencakup lebih dari satu peubah prediktor pada level terendah demikian pula pada level yang lebih tinggi. Menurut Hox (2018) model regresi multilevel untuk level 1 pada unit level 2 adalah sebagai berikut:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij}, \quad (1)$$

dimana

$Y_{ij}$  : Variabel respon pada level 1,

$\beta_{0j}$  : Intersep pada level 1,

$\beta_1$  : Koefisien regresi pada level 1,

$X_{ij}$  : Variabel prediktor pada level 1, dan

$e_{ij}$  : Galat pada level 1.

Indeks  $i$  menyatakan unit level 1 yang tersarang dalam level 2 dan indeks  $j$  menyatakan unit level 2. Persamaan pada level 2 intersep pada level 1  $\beta_{0j}$  menjadi peubah respon pada level 2 sebagai berikut:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}, \quad (2)$$

dimana

$\gamma_{00}$  : Intersep pada level 2,

$\gamma_{01}$  : Koefisien regresi pada level 2,

$Z_j$  : Variabel prediktor pada level 2,

$u_{0j}$  : Galat pada level 2.

Dengan mensubstitusikan persamaan (2) ke persamaan (1) diperoleh model regresi multilevel (dua level) sebagai berikut:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} + \beta_1 X_{ij} + e_{ij}, \quad (3)$$

Berikut merupakan model multilevel survival:

$$h(t|X, u_j, e_{ij}) = h_0(t) e^{(\beta_{0j} + \beta_1 X_{ij} + e_{ij})} \quad (4)$$

dengan

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_j,$$

Metode estimasi parameter dalam model ada dua pendekatan yaitu pendekatan klasik dan pendekatan Bayesian. Pada pendekatan klasik sepenuhnya mengandalkan proses inferensi pada data sampel yang diambil dari populasi. Sedangkan pendekatan Bayesian, disamping memanfaatkan data sampel juga memperhitungkan suatu distribusi awal yang disebut distribusi *prior* (Rahmi, 2016). Teorema Bayesian didasarkan pada pada distribusi posterior yang merupakan perpaduan antara distribusi prior dan data observasi yang digunakan untuk menyusun fungsi *likelihood* (Box & Tiao, 1973).

Metode *Markov Chain Monte Carlo* (MCMC) merupakan metode pendekatan untuk inferensi Bayesian. Menurut Casella (2009) MCMC digunakan untuk membangkitkan peubah-peubah acak yang didasarkan pada rantai markov. Algoritma *Gibbs Sampling* merupakan salah satu algoritma yang terdapat dalam MCMC. Algoritma *Gibbs Sampling* digunakan apabila terdapat lebih dari satu parameter yang tidak diketahui. Untuk melakukan perhitungan digunakan program WinBUGS. WinBUGS adalah paket program yang dirancang untuk pemodelan data dengan pendekatan Bayesian menggunakan implementasi *Markov Chain Monte Carlo* (MCMC). Pengolahan data menggunakan WinBUGS dilakukan dengan tahap yang pertama yaitu spesifikasi model, inisialisasi, menentukan iterasi MCMC sebanyak 10.000 kali untuk membangkitkan rantai markov (Lukitasari *et al.*, 2015).

Beberapa penelitian yang telah dilakukan sebelumnya oleh Zahnd *et al.* (2019) Analisis Multilevel dalam pengendalian kanker di pedesaan. Mustafafi (2015) juga melakukan pemodelan multilevel *survival* yang diterapkan pada data infeksi berulang penyakit *Chronic Granulomatous Disease* (CGD) namun dengan pendekatan *hierarchical likelihood*. Penelitian berikutnya oleh Octaviana (2017) penelitian ini

memodelkan status bekerja ibu rumah tangga menggunakan Model Multilevel dengan respon biner.

Ketiga penelitian tersebut menggunakan metode *Maximum Likelihood Estimation* (MLE) dalam mengestimasi parameter. Tetapi teknik ini tidak bisa digunakan jika distribusi datanya tidak diketahui. Dari hal tersebut maka pada penelitian ini menggunakan Bayesian untuk mencari estimasi parameternya. Pada metode Bayesian parameter populasi dipandang sebagai variabel yang memiliki distribusi awal (prior). Sebelum menarik sampel dari suatu populasi terkadang diperoleh informasi mengenai parameter yang akan diestimasi. Informasi ini kemudian digabungkan dengan informasi dari sampel untuk digunakan dalam mengestimasi parameter populasi sehingga metode Bayesian lebih menjanjikan karena ada informasi tambahan untuk menyimpulkan karakteristik populasi. Keuntungan lain dari penggunaan pendekatan Bayesian yaitu mampu menyediakan inferensia dan estimasi yang stabil meski pada situasi dimana ukuran sampel sangat kecil.

Berdasarkan permasalahan di atas, maka tujuan penelitian ini adalah memperoleh estimasi parameter model multilevel *survival* menggunakan metode Bayesian *markov chain monte carlo*, memperoleh model multilevel *survival* pada penderita Demam Berdarah *Dengue* (DBD), dan mengetahui variabel yang mempengaruhi waktu kelangsungan hidup penderita Demam Berdarah *Dengue* (DBD).

**METODE**

Data penelitian yang digunakan adalah data sekunder dari RSUD Tugurejo pada tahun 2015 tentang pasien rawat inap penyakit Demam Berdarah *Dengue* (DBD). Terdapat 90 pasien dengan 3 pasien meninggal (tidak tersensor) dan 87 pasien sembuh (tersensor).

Variabel yang digunakan dalam penelitian ini adalah variabel respon yang merupakan waktu *survival*. Sedangkan variabel prediktor pada level 1 meliputi:  $X_1$  = Usia,  $X_2$  = jenis kelamin (1=laki-laki, 2=perempuan),  $X_3$  = kejadian pendarahan (1=terjadi pendarahan, 2=tidak terjadi pendarahan),  $X_4$  = jumlah trombosit,  $X_5$  = kadar hematokrit. Variabel prediktor pada level 2 meliputi:  $Z_1$  = curah hujan,  $Z_2$  = jumlah pasien tiap bulan.

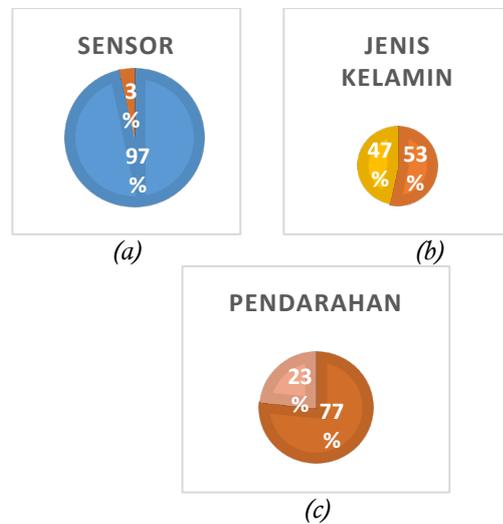
Metode penelitian yang digunakan dalam penelitian ini adalah analisis *survival* dengan menggunakan pendekatan Bayesian *markov chain monte carlo* dengan algoritma *Gibbs*

*Sampling*. Adapun langkah-langkah dalam melakukan penelitian adalah sebagai berikut: (1) Melakukan pengecekan asumsi *Proportional Hazard* untuk setiap variabel prediktor. (2) Menentukan distribusi pada data waktu *survival* pasien DBD. (3) Mengestimasi parameter pada distribusi data waktu *survival*. (4) Menentukan fungsi *survival* dan fungsi *hazard*. (5) Mengestimasi parameter pada model multilevel *survival*. (6) Membantu model multilevel *survival* dan menentukan faktor-faktor yang mempengaruhi waktu *survival* pada pasien DBD

**HASIL DAN PEMBAHASAN**

**1. Karakteristik Pasien Demam Berdarah *Dengue* (DBD)**

Terdapat 90 sampel data dengan 3 pasien meninggal (tidak tersensor) dan 87 pasien sembuh (tersensor). Dari 90 pasien, sebanyak 53% (48 pasien) berjenis kelamin laki-laki dan 47%(42 pasien) berjenis kelamin perempuan. Sebanyak 77% pasien mengalami pendarahan dan 23% tidak mengalami pendarahan. Berikut akan ditampilkan pada Gambar 1.



Gambar 1. *Pie Chart* variabel (a) sensor, (b) jenis kelamin, (c) kejadian pendarahan

Berdasarkan data pasien maka didapat dilihat deskriptif pasien Demam Berdarah *Dengue* (DBD) yang ditampilkan pada Tabel 1 sebagai berikut:

Tabel 1. Deskripsi Pasien DBD di RSUD Tugurejo Provinsi Jawa Tengah

Variabel	Mean	Std. Dev.	Min	Max
Lifetime	5.42	1.628	3	10
Usia	11,37	13.131	1	69
Trombosit	142106.67	128245.331	9200	497000
Hematrokrit	38.67	5.212	29	54

Tabel 1 menunjukkan bahwa waktu *survival* pasien atau lama waktu rawat inap pasien (hingga diperbolehkan pulang karena sembuh) paling cepat selama 3 hari dan rawat inap paling lama selama 10 hari dengan rata-rata lama rawat inap pasien DBD sekitar 5 hari. Sedangkan rata-rata usia pasien DBD sekitar 11 tahun dengan usia paling kecil 1 tahun dan usia paling tua 69 tahun. Rata-rata jumlah trombosit sekitar 142107/ml dan nilai minimum sebesar 9200/ml yang dikategorikan sangat rendah karena sangat jauh dari angka 100000/ml. Rata-rata kadar hematrokrit adalah sekitar 39%, sedangkan nilai maksimum kadar hematrokrit sebesar 54% yang bisa dikatakan tinggi karena >46%.

2. Distribusi *Lifetime* pada data

Pengujian distribusi *Lifetime* digunakan untuk mengetahui distribusi data yang digunakan dalam penelitian. Untuk menguji distribusi data digunakan *scatter plot* dan nilai *Anderson Darling*. Sampel data pasien DBD diuji dengan empat distribusi yang biasanya digunakan dalam analisis *survival* yaitu distribusi Normal, Lognormal, Eksponensial, dan Weibull. Berikut merupakan hasil *scatter plot*. Berikut ditampilkan pada Gambar 2.

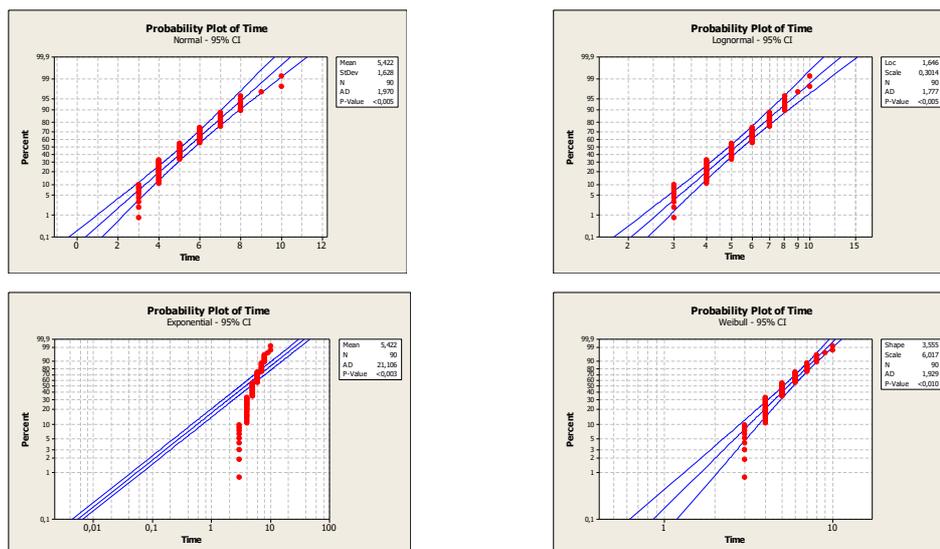
Berdasarkan Gambar 2 dapat dilihat bahwa plot yang paling mendekati garis adalah distribusi Lognormal. Artinya data *lifetime* pada pasien DBD berdistribusi Lognormal.

Pengujian selanjutnya adalah dengan melihat nilai *Anderson Darling*. Berikut merupakan nilai *Anderson Darling* yang ditampilkan pada Tabel 2.

Tabel 2. Anderson Darling

Distribusi	Anderson Darling
Normal	1,970
Lognormal	1,777
Exponential	21,106
Weibull	1,929

Untuk menguji distribusi data, digunakan nilai *Anderson Darling* paling kecil. Berdasarkan Tabel 2 dapat dilihat bahwa yang memiliki nilai *Anderson Darling* paling kecil adalah distribusi Lognormal. Sehingga dapat diasumsikan bahwa data pasien DBD di RSUD Tugurejo Provinsi Jawa Tengah berdistribusi Lognormal.



Gambar 2. Scatter plot probabilitas *lifetime* data pasien DBD

**3. Pengecekan Asumsi Proportional Hazard (PH)**

Asumsi *proportional hazard* merupakan asumsi yang harus dipenuhi dalam analisis *survival*. Dalam pengujian asumsi PH digunakan uji *global test* atau *Goodness of Fit* (GOF). Berikut merupakan nilai *p-value* yang ditampilkan pada Tabel 3.

Tabel 3. Nilai *p-value* Global Test

Peubah	<i>p-value</i>
Usia ( $X_1$ )	299,607
Jenis Kelamin ( $X_2$ )	6,352
Kejadian Pendarahan ( $X_3$ )	7,050
Jumlah Trombosit ( $X_4$ )	466,333
Kadar Hematrokrit ( $X_5$ )	163,983

Berdasarkan Tabel 3 menunjukkan bahwa dari hasil *global test* untuk tiap peubah, masing-masing memiliki *P-value* >  $\alpha$ , dengan  $\alpha = 0,05$ , maka dapat disimpulkan bahwa asumsi *proportional hazard* terpenuhi.

**4. Estimasi Parameter Distribusi Penduga**

Estimasi parameter model *survival* pada penelitian ini menggunakan metode Bayesian *markov chain monte carlo* dengan algoritma *Gibbs Sampling*. Langkah-langkah estimasi parameter model pada analisis *survival* adalah sebagai berikut:

- a. Menentukan fungsi kepadatan peluang dari *lifetime* penderita DBD  
*Lifetime* pada data penderita DBD mengikuti distribusi lognormal dengan parameter  $\mu$  dan  $\sigma^2$ , bentuk distribusinya adalah:

$$f(t|\mu, \sigma^2) = \frac{1}{\sigma t \sqrt{2\pi}} \left[ -\frac{1}{2} \left( \frac{\ln t - \mu}{\sigma} \right)^2 \right], \sigma^2 > 0, t > 0 \quad (5)$$

- b. Menentukan fungsi *likelihood*  
 Jika  $t_1, t_2, \dots, t_n$  adalah sampel acak berdistribusi lognormal dengan fungsi kepadatan peluang  $f(t|\mu, \sigma)$  maka fungsi *likelihood* nya didefinisikan dengan:

$$L(t_i|\mu, \sigma) = \prod_{i=1}^n f(t_i|\mu, \sigma) = \prod_{i=1}^n \left( \frac{1}{\sigma t_i \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln t_i - \mu}{\sigma} \right)^2 \right] \right) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n \frac{1}{t_i} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln(t_i) - \mu)^2 \right] \quad (6)$$

- c. Menentukan distribusi *prior*  
 Distribusi *prior* ditentukan berdasarkan ruang parameternya. Dalam penelian ini, *lifetime* berdistribusi

lognormal dengan parameter  $\mu$  dan  $\sigma^2$  dengan  $-\infty < \mu < \infty$  dan  $\sigma > 0$ .

Digunakan distribusi *prior* Normal untuk  $\mu$  dengan parameter  $\theta$  dan  $\tau$ , fungsi kepadatan peluangnya sebagai berikut:

$$f(\mu|\theta, \tau) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left( -\frac{(\mu - \theta)^2}{2\tau^2} \right) \quad (7)$$

dimana  $-\infty < \mu < \infty, -\infty < \theta < \infty, \tau^2 > 0$ .

Digunakan distribusi *prior* Invers Gamma untuk  $\sigma^2$  dengan parameter  $\alpha$  dan  $\beta$ , fungsi kepadatan peluangnya sebagai berikut:

$$f(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-\alpha-1} \exp \left( -\frac{\beta}{\sigma} \right) \quad (8)$$

dimana  $\tau > 0, \alpha \geq 0, \beta \geq 0$ .

- d. Membentuk distribusi *posterior*  
 Setelah menentukan fungsi *likelihood* dan distribusi *prior* maka dapat dicari distribusi *posterior*. Distribusi *posterior* dapat dinyatakan sebagai berikut:

$$f(\mu, \sigma^2|t_i) \propto L(t_i|\mu, \sigma^2) p(\mu) p(\sigma^2) = \left( \frac{1}{\sigma t \sqrt{2\pi}} \right)^n \prod_{i=1}^n \frac{1}{t_i} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \ln(t_i - \mu)^2 \right] \times \frac{1}{\sqrt{2\pi\tau^2}} \exp \left( -\frac{(\mu - \theta)^2}{2\tau^2} \right) \times \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-\alpha-1} \exp \left( -\frac{\beta}{\sigma} \right) \quad (9)$$

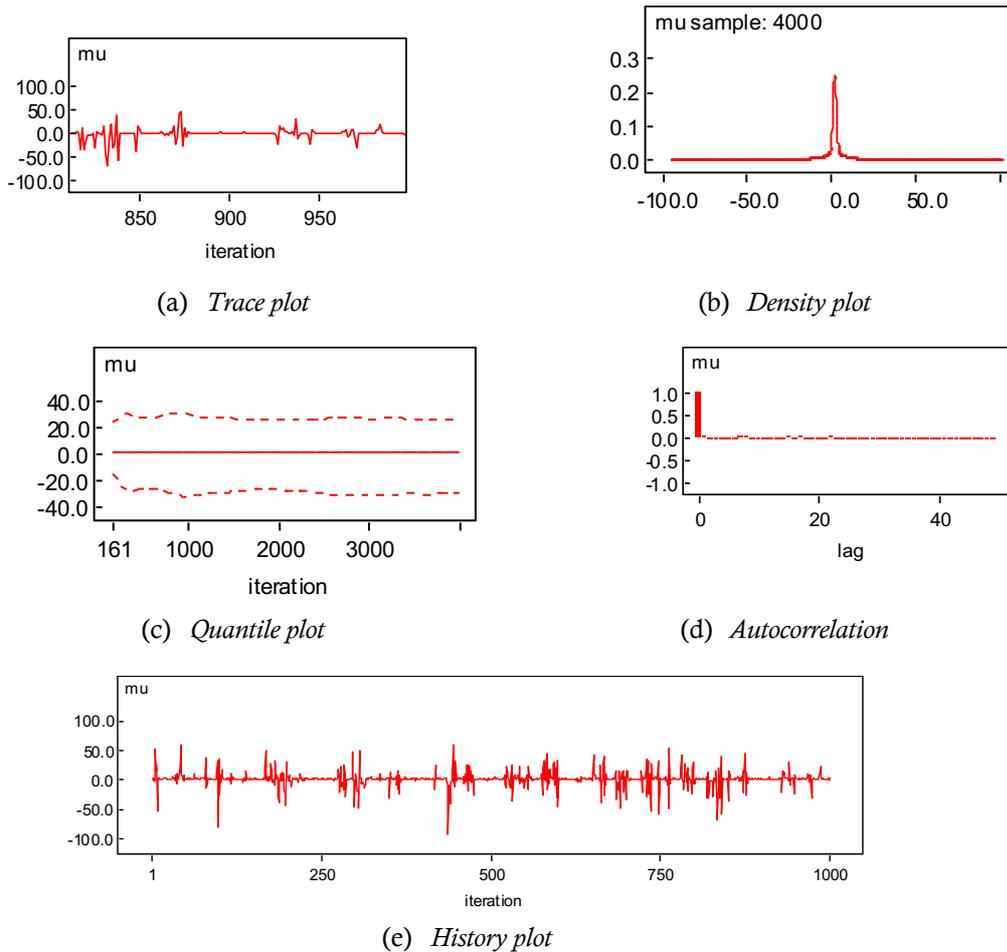
Persamaan pada distribusi *posterior* sangat kompleks sehingga sulit untuk memecahkannya, maka untuk melakukan iterasi menggunakan *Markov Chain Monte Carlo* (MCMC) dengan algoritma *Gibbs sampling*.

- e. Mengestimasi parameter dengan MCMC algoritma *Gibbs Sampling* dengan bantuan program WinBUGS. Langkah-langkahnya sebagai berikut:
  - Menentukan nilai awal dari  $\{\mu^{(0)}, \sigma^{2(0)}\}$
  - Bangkitkan Nilai  $\mu^{(1)}$  berasal dari  $f(\mu|\sigma^{2(0)})$
  - Nilai  $\sigma^{2(1)}$  berasal dari  $f(\sigma^2|\mu^{(1)})$  Sehingga diperoleh parameter baru  $\{\mu^{(1)}, \sigma^{2(1)}\}$
  - Gunakan  $\{\mu^{(1)}, \sigma^{2(1)}\}$  untuk nilai awal proses selanjutnya sampai  $k$  iterasi yang diinginkan.

- Periksa konvergensi dengan melihat *trace plot*, *history plot*, *autocorrelation*, *quartile plot*, dan *density posterior*.
- Mendapatkan nilai mean, median, standar deviasi dari distribusi *posterior*.

Pada Gambar 3 berikut merupakan output dari program WinBUGS.

(d) plot nilai-nilai sampel terhadap lognya mendekati nol. Hal ini menunjukkan



Gambar 3. Diagnostic plot parameter lognormal

Gambar 3 (a) mengilustrasikan suatu *trace plot* yang menunjukkan bahwa proses rantai markov atau *burn-in* sudah mencapai pada sebaran targetnya atau dapat diartikan proses *burn-in* sudah selesai dengan kata lain sudah mencapai kekonvergenan. Gambar 3 (b) menunjukkan *density plot* dengan kurva yang memiliki titik tengah sehingga dapat dikatakan bahwa model sudah konvergen. *Quantile plot* pada Gambar 3 (c) menunjukkan bahwa posisi plot terletak diantara batas atas dan batas bawah. *Autocorrelation* pada Gambar 3

bahwa korelasi pada rantai markov lemah, sehingga algoritma sudah berada dalam distribusi target. Pada Gambar 3 (e) adalah *history plot* dapat dilihat bahwa plot terlihat rapat dan merespon semua variabel sehingga menunjukkan bahwa model sudah konvergen.

Selanjutnya adalah mencari nilai estimasi tiap parameter berikut ditampilkan pada Tabel 4 berikut:

Tabel 4. Estimasi parameter distribusi

lognormal				
Paramet er	Mean	2,5%	Median	97,5%
$\mu$	1,531	-29,62	1,796	27,32
$\sigma$	45,0	0,04752	3,004	221,0

**5. Fungsi Survival dan Fungsi Hazard**

Fungsi kepadatan peluang dari distribusi lognormal adalah:

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln t - \mu}{\sigma} \right)^2 \right], \quad \sigma > 0, t > 0 \quad (10)$$

Sedangkan fungsi distribusi kumulatif  $F(t)$  memiliki persamaan sebagai berikut:

$$F(t) = \left( \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left( \frac{\ln t - \mu}{\sigma \sqrt{2}} \right) \right) \quad (11)$$

Untuk fungsi survivalnya sebagai berikut:

$$S(t) = 1 - F(t) = 1 - \left( \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left( \frac{\ln t - \mu}{\sigma \sqrt{2}} \right) \right) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \frac{\ln t - \mu}{\sigma \sqrt{2}} \right) \quad (12)$$

Fungsi hazardnya adalah:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{\sigma t \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln t - \mu}{\sigma} \right)^2 \right]}{\frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \frac{\ln t - \mu}{\sigma \sqrt{2}} \right)} = \frac{\sqrt{2} \exp \left[ -\frac{1}{2} \left( \frac{\ln t - \mu}{\sigma} \right)^2 \right]}{\sigma \sqrt{\pi} \left( 1 - \operatorname{erf} \left( \frac{\ln t - \mu}{\sigma \sqrt{2}} \right) \right)} \quad (13)$$

Untuk menghitung nilai fungsi *survival* dan fungsi *hazard* maka digunakan estimasi parameter dari distribusi lognormal. Berdasarkan persamaan (12) dan (13) maka didapat Tabel 5 sebagai berikut:

Tabel 5. Estimasi Fungsi Survival dan Fungsi Hazard

t	S(t)	h(t)
1	0,51357	0,017257
2	0,507427	0,008736
3	0,503833	0,005867
4	0,51283	0,004422
5	0,400305	0,003552

6	0,4976688	0,00297
7	0,496322	0,002552
8	0,495138	0,002239
9	0,494094	0,001994
10	0,49316	0,001798

**6. Estimasi parameter model multilevel survival**

Dalam penelitian ini *lifetime* berdistribusi Lognormal dengan parameter  $(\mu, \sigma^2)$ . Model ditentukan sebagai berikut:

$$t_i \sim \operatorname{Lognormal}(\mu_i, \sigma^2)$$

Dimana  $\mu_i = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$

langkah-langkah untuk mengestimasi parameter model multilevel adalah sebagai berikut:

- a. Menentukan fungsi kepadatan peluang dari *lifetime* penderita DBD

*Lifetime* pada data penderita DBD mengikuti distribusi lognormal dengan parameter  $\mu$  dan  $\sigma^2$ , bentuk distribusinya adalah:

$$f(t|\mu, \sigma^2) = \frac{1}{\sigma t \sqrt{2\pi}} \left[ -\frac{1}{2} \left( \frac{\ln t - \mu}{\sigma} \right)^2 \right], \quad \sigma^2 > 0, t > 0 \quad (14)$$

- b. Menentukan fungsi *likelihood*

Jika  $t_1, t_2, \dots, t_n$  adalah sampel acak berdistribusi lognormal dengan fungsi kepadatan peluang  $f(t|\mu, \sigma)$  maka fungsi *likelihood* nya didefinisikan dengan:

$$L(t_i|\mu, \sigma) = \prod_{i=1}^n f(t_i|\mu, \sigma) = \prod_{i=1}^n \left( \frac{1}{\sigma t_i \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln t_i - \mu}{\sigma} \right)^2 \right] \right) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n \frac{1}{t_i}$$

$$\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln(t_i) - \mu)^2 \right] \quad (15)$$

- c. Menentukan distribusi *prior*

Distribusi *prior* ditentukan berdasarkan ruang parameternya. Dalam penelian ini, *lifetime* berdistribusi lognormal dengan parameter  $\mu$  dan  $\sigma^2$  dengan  $-\infty < \mu < \infty$  dan  $\sigma > 0$ .

Digunakan distribusi *prior* Normal untuk  $\mu$  dengan parameter  $\theta$  dan  $\tau$ , fungsi kepadatan peluangnya sebagai berikut:

$$f(\mu|\theta, \tau) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\mu - \theta)^2}{2\tau^2}\right) \quad (16)$$

dimana  $-\infty < \mu < \infty, -\infty < \theta < \infty, \tau^2 > 0$ .

Digunakan distribusi *prior* Invers Gamma untuk  $\sigma^2$  dengan parameter  $\alpha$  dan  $\beta$ , fungsi kepadatan peluangnya sebagai berikut:

$$f(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma}\right) \quad (17)$$

dimana  $\tau > 0, \alpha \geq 0, \beta \geq 0$ .

d. Membentuk distribusi *posterior*

Setelah menentukan fungsi *likelihood* dan distribusi *prior* maka dapat dicari distribusi *posterior*. Distribusi *posterior* dapat dinyatakan sebagai berikut:

$$\begin{aligned} f(\mu, \sigma^2|t_i) &\propto L(t_i|\mu, \sigma^2)p(\mu)p(\sigma^2) \\ &\propto \left(\frac{1}{\sigma t\sqrt{2\pi}}\right)^n \prod_{i=1}^n \frac{1}{t_i} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \ln(t_i - \mu)^2\right] \\ &\quad \times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\mu - \theta)^2}{2\tau^2}\right) \\ &\quad \times \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma}\right) \quad (18) \end{aligned}$$

Persamaan pada distribusi *posterior* sangat kompleks sehingga sulit untuk memecahkannya, maka untuk melakukan iterasi menggunakan *Markov Chain Monte Carlo* (MCMC) dengan algoritma *Gibbs sampling*.

e. Mengestimasi parameter dengan MCMC dengan algoritma *Gibbs Sampling*

Langkah-langkahnya sebagai berikut:

- Menentukan nilai awal dari  $\theta^{(0)} =$

$$\{\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_n^{(0)}, \sigma^{(0)}\}$$

- Bangkitkan

Nilai  $\beta_1^{(1)}$  berasal dari  $f(\beta_1|\beta_2^{(0)}, \dots, \beta_n^{(0)}, \sigma^{(0)})$

Nilai  $\beta_2^{(1)}$  berasal dari  $f(\beta_2|\beta_1^{(1)}, \beta_3^{(1)}, \dots, \beta_n^{(0)}, \sigma^{(0)})$

⋮

Nilai  $\beta_n^{(1)}$  berasal dari  $f(\beta_n|\beta_1^{(1)}, \dots, \beta_{n-1}^{(1)}, \sigma^{(0)})$

Nilai  $\sigma^{(1)}$  berasal dari  $\sigma(\beta_1^{(1)}, \dots, \beta_n^{(1)})$

Sehingga diperoleh parameter baru

$$\{\beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_n^{(1)}, \sigma^{(1)}\}$$

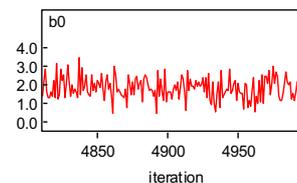
- Gunakan  $\{\beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_n^{(1)}, \sigma^{(1)}\}$  untuk nilai awal proses selanjutnya sampai  $k$  iterasi yang diinginkan.
- Periksa konvergensi dengan melihat *trace plot*, *history plot*, *autocorrelation*, *quartile plot*, dan *density posterior*.
- Mendapatkan nilai mean, median, standar deviasi dari distribusi *posterior*.
- Membentuk model *proportional hazard* dari hasil taksiran parameter

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (18)$$

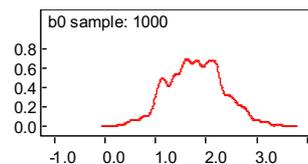
dimana nilai  $\beta_1, \beta_2, \dots, \beta_n$  adalah nilai mean *posterior* dari proses *Markov Chain Monte Carlo* (MCMC) yang telah konvergen.

Estimasi parameter pada penelitian ini menggunakan software WinBUGS. Iterasi dilakukan sampai mencapai konvergen yang dapat diketahui dari *trace plot*, *density plot*, *quantile plot*, *autocorrelation* dan *history plot*. Pada proses estimasi parameter ini dilakukan sebanyak 5000 iterasi menggunakan MCMC dengan algoritma *Gibbs Sampling*. Dari iterasi sebanyak 5000 iterasi menunjukkan bahwa proses estimasi sudah mencapai kondisi konvergen.

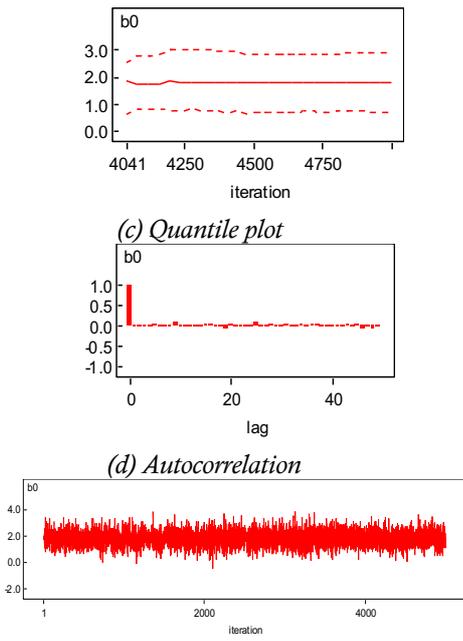
Berikut merupakan output dari estimasi parameter menggunakan software WinBUGS yang ditampilkan pada Gambar 4.



(a) Trace plot



(b) Density plot



(e) History plot  
 Gambar 4. *Diagnostic plot* parameter multilevel survival

Berdasarkan data pasien demam berdarah di RSUD Tugurejo Jawa Tengah yang digunakan dalam penelitian ini, diperoleh hasil estimasi parameter masing-masing variabel yang disajikan pada Tabel 6.

7. **Pemodelan Multilevel Survival**

Tabel 6. Hasil taksiran parameter

Model	Parameter	Nilai Estimasi	Interval Kepercayaan	
			2,5%	97,5%
Level 1	$\beta_{Usia}$	-3,929E-4	-0,002804	0,002318
	$\beta_{Jenis\ Kelamin}$	0,0412	0,04004	0,1167
	$\beta_{Kejadian\ Pendarahan}$	0,2592	0,139	0,3794
	$\beta_{Jumlah\ Trombosit}$	-5,137E-7	-9,381E-7	-7,028E-8
	$\beta_{Kadar\ Hematokrit}$	0,002516	-0,001573	0,008687
Level 2	$\gamma_{00}$	1,799	0,7406	2,917
	$\gamma_{Curah\ Hujan}$	-5,867E-4	-9,676E-4	-2,309E-4
	$\gamma_{Jumlah\ Pasien}$	0,03977	0,01842	0,05226

Dari Tabel 6 diketahui nilai taksiran untuk masing-masing parameter kovariat. Berdasarkan Persamaan (4) diperoleh hasil sebagai berikut:

$$\begin{aligned}
 h(t|X, u_j, e_{ij}) = & h_0(t) \exp(1,799 \\
 & - 5,867 \cdot 10^{-4} Z_1 \\
 & + 0,03977 Z_2 \\
 & - 3,929 \cdot 10^{-4} X_1 \\
 & + 0,0412 X_2 + 0,2592 X_3 \\
 & - 5,137 \cdot 10^{-7} X_4 \\
 & + 0,002516 X_5)
 \end{aligned}$$

Berdasarkan Tabel 6 nilai estimasi parameter untuk variabel Usia ( $X_1$ ) dan variabel Kadar Hematokrit ( $X_5$ ) memiliki nilai parameter dengan interval kepercayaan pada selang 2,5% sampai 97,5% melewati nilai 0, sehingga variabel tersebut tidak masuk dalam model. Dapat disimpulkan bahwa variabel jenis kelamin, terjadinya pendarahan, jumlah trombosit, curah hujan dan jumlah pasien per bulan berpengaruh secara signifikan terhadap lama rawat inap pasien DBD. Jadi kovariat selain kejadian dan jumlah pasien perbulan dikeluarkan dari model, sehingga diperoleh model sebagai berikut:

$$\begin{aligned}
 h(t|X, u_j, e_{ij}) = & h_0(t) \exp(1,799 \\
 & - 5,867 \cdot 10^{-4} Z_1 \\
 & + 0,03977 Z_2 + 0,0412 X_2 \\
 & + 0,2592 X_3 \\
 & - 5,137 \cdot 10^{-7} X_4)
 \end{aligned}$$

Hasil analisis dari model diatas menunjukkan bahwa kejadian pendarah memiliki pengaruh yang paling besar dari semua variabel yang berpengaruh terhadap waktu kelangsungan hidup penderita demam berdarah karena dalam model terlihat bahwa nilai koefisien dari variabel kejadian pendarahan ( $X_3$ ) memiliki nilai koefisien yang paling besar dibandingkan dengan variabel yang lainnya yaitu sebesar 0,2592.

**8. Aplikasi Persamaan *Proportional Hazard Multilevel Survival***

Dari hasil analisis yang telah dilakukan maka dapat diaplikasikan pada kasus pasien DBD. Misalkan seorang pasien DBD berjenis kelamin laki-laki, tidak mengalami pendarahan, dengan jumlah trombosit  $109000/mm^3$ . Pada bulan tersebut terdapat 15 pasien yang dirawat dan saat bulan tersebut mengalami curah hujan sebesar 238mm. Akan dihitung *hazard rate* setelah 6 hari pengobatan di rumah sakit.

Diketahui:  $h_0(t) = 0,00297$

$$s_0(t) = 0,4976688$$

Jadi *hazard rate* adalah:

$$h(t|X, u_j, e_{ij}) = h_0(t) \exp(1,799 - 5,867 \cdot 10^{-4} Z_1 + 0,03977 Z_2 + 0,0412 X_2 + 0,2592 X_3 - 5,137 \cdot 10^{-7} X_4)$$

$$h(t) = 0,00297 \exp(1,799 - 5,867 \cdot 10^{-4} \times 238 + 0,03977 \times 15 + 0,0412 \times 0 + 0,2592 \times 0 - 5,137 \cdot 10^{-7} \times 109000)$$

$$h(t) = 0,00297 \exp(2,199922)$$

$$h(t) = 0,026802$$

dan *survival rate* adalah:

$$S(t) = s_0(t)^y$$

dengan

$$y = \exp(1,799 - 5,867 \cdot 10^{-4} Z_1 + 0,03977 Z_2 + 0,0412 X_2 + 0,2592 X_3 - 5,137 \cdot 10^{-7} X_4)$$

$$y = \exp(1,799 - 5,867 \cdot 10^{-4} \times 238 + 0,03977 \times 15 + 0,0412 \times 0 + 0,2592 \times 0 - 5,137 \cdot 10^{-7} \times 109000)$$

$$y = \exp(2,199922)$$

$$S(t) = 0,4976688^{\exp(2,199922)}$$

$$S(t) = 0,21542$$

Jadi peluang mengalami kematian pada pasien DBD berjenis kelamin laki-laki yang tidak mengalami pendarahan, dengan jumlah trombosit  $109000/mm^3$ . Pada bulan tersebut terdapat 15 pasien yang dirawat dan saat bulan tersebut mengalami curah hujan sebesar 238mm yang dirawat selama 6 hari adalah sebesar 2,8% dan peluang bertahan hidup sebesar 21,5%.

**PENUTUP**

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan maka diperoleh estimasi parameter dari Distribusi Lognormal dengan pendekatan Bayesian yaitu  $\mu = 1,531$  dan  $\sigma = 45,0$ . Model *hazard proportional multilevel survival* yang diterapkan pada data Demam Berdarah *Dengue* (DBD) adalah sebagai berikut:

$$h(t|X, u_j, e_{ij}) = h_0(t) \exp(1,799 - 5,867 \cdot 10^{-4} Z_1 + 0,03977 Z_2 + 0,0412 X_2 + 0,2592 X_3 - 5,137 \cdot 10^{-7} X_4)$$

dengan  $z_1$  adalah curah hujan,  $z_2$  adalah jumlah pasien perbulan,  $x_2$  adalah jenis kelamin,  $x_3$  adalah kejadian pendarahan, dan  $x_4$  adalah jumlah trombosit.

Dapat disimpulkan bahwa faktor yang berpengaruh terhadap waktu ketahanan hidup pasien DBD di RSUD Tugurejo pada tahun 2015 adalah curah hujan, jumlah pasien, jenis kelamin, kejadian pendarahan dan jumlah trombosit.

Pada hasil penelitian terdapat faktor yang tidak mempengaruhi waktu kelangsungan penderita DBD. Maka diharapkan untuk bisa mengembangkan metode analisis *survival* lain seperti, *non proportional hazard*, model *failure*, model *mixture*, model *frailty* dan model lain dan pada penelitian ini belum mencakup semua variable yang mungkin akan berpengaruh terhadap waktu kelangsungan hidup seperti efek spasial, karena pada efek spasial mungkin akan sangat berpengaruh terutama pasien yang hidup di tempat yang lembab. Maka diharapkan pada penelitian selanjutnya bisa menambahkan efek spasial pada penelitiannya.

**UCAPAN TERIMAKASIH**

Terimakasih kepada keluarga yang telah memberikan dukungan lahir serta batin. Serta pihak-pihak yang membantu dan mendukung selama pelaksanaan penelitian ini.

**DAFTAR PUSTAKA**

Antonio, K., Frees, E. W., & Valdez, E. A. 2010. A Multilevel Analysis of Intercompany Claim Count. *ASTIN Bulletin*, 40(1): 151-177.

Box, G. E. P., & Tiao, G. C. 1973. *Bayesian Inference in Statistical Analysis*. In *Bayesian Inference in Statistical Analysis*. New York: John Willey and Sons, Inc.

- Casella, G. & Berger, R.L. 2009. *Statistical Inference*. USA: Thomson Learning, Duxbury.
- Ha, L. D., & Lee, Y. 2005. Multilevel Mixed Linear Model for Survival Data. *Lifetime Data Analysis*, 11(1): 131-142.
- Harlan, J. 2016. *Analisis Multilevel*. Jakarta: Gunadarma.
- Hox, J. J., Moerbeek, M., & Schoot, R. van de. 2018. *Multilevel Analysis Techniques and Applications*. New York: Routledge.
- Khan, H. R., & Shaw, J. E. H. 2011. Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data. *Journal of Data Science*, 9(1): 93-110.
- Lee, E. T., & Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis*. New York: A John Wiley & Sons, Inc.
- Luh, N., & Ayu, P. 2016. Analisis Regresi Multilevel dalam Menentukan Variabel Determinan Nilai Ujian Akhir Nasional Siswa. *Jurnal Matematika*. 6(1): 15-21.
- Lukitasari, D., Setiawan, A., & Sasangko, L. R. 2015. Bayesian Survival Analysis Untuk Mengestimasi Parameter Model Weibull-Regression Pada Kasus Ketahanan Hidup Pasien Penderita Jantung Koroner. *d'CARTESIAN*. 4(1): 26-33.
- Lukociene, O., & Vermunt, J. K. 2008. A Comparison of Multilevel Logistic Regression Models with Parametric and Nonparametric Random Intercept. *Preprint Submitted to Elsevier Science*.
- Mustafafi, R. D. 2015. *Analysis Data Multilevel Survival menggunakan Metode Hierarchical Likelihood*. Thesis. Universitas Gadjah Mada.
- Octaviana, F. A. 2017. *Pemodelan Status Bekerja Ibu Rumah Tangga menggunakan Model Multilevel dengan Respon Biner*. Sepuluh Nopember Institute of Technology Surabaya.
- Rahmi, S. 2016. *Pemodelan Survival Spasial dengan Pendekatan Bayesian*. Makassar: Universitas Hasanuddin.
- Raman, R. & Hedekar, D. 2005. A Mixed Effects Regression Model for Three Level Ordinary Respon Data. *Statistics in Medicine*, 24(21): 3331-3345.
- Silva, P. J. 2007. An Introduction to Multilevel Modeling for Research on the Psychology of Art and Creativity. *Empirical Studies of Art*. 25(1): 1-20.
- Yau, K. K. W. 2001. Multilevel Model for Survival Analysis with Random Effects. *Biometrics*. 57(1):96-102.