



**APLIKASI MODEL REGRESI PH COX UNTUK
ANALISIS TAHAN HIDUP MENGGUNAKAN
PENDEKATAN BAYESIAN**

SKRIPSI

**diajukan untuk memenuhi salah satu syarat
untuk memperoleh gelar Sarjana Sains
Program Studi Matematika**

oleh

Yunita Dwi Ayu Ningtias

4111414018

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2019

PERNYATAAN KEASLIAN TULISAN

Saya menyatakan bahwa skripsi ini bebas plagiat, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti terdapat plagiat dalam Skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan yang berlaku.

Semarang, Mei 2019

Peneliti,



Yunita Dwi Ayu Ningtias

4111414018

PENGESAHAN

Skripsi yang berjudul

Aplikasi Model Regresi PH Cox untuk Analisis Tahan Hidup Menggunakan Pendekatan Bayesian

Disusun oleh

Yunita Dwi Ayu Ningtias

4111414018

Telah dipertahankan dihadapan Panitia Penguji Skripsi FMIPA Universitas Negeri Semarang pada hari Kamis, tanggal 2 Mei 2019.

Panitia :

Ketua



Prof. Dr. Sudarmin, M.Si

NIP. 196601231992031003

Sekretaris

Drs Arief Agoestanto, M.Si

NIP. 196807221993031005

Ketua Penguji

Dr. Scolastika Mariani, M.Si

NIP. 19650201991022001

Anggota Penguji/

Pembimbing I

Dr. Nur Karomah Dwidayati, M.Si

NIP. 196605041990022001

Anggota Penguji/

Pembimbing II

Drs Arief Agoestanto, M.Si

NIP. 196807221993031005

MOTTO

“Memilihlah dengan tanpa penyesalan.”

(Mary Anne Radmacher)

“Tiada keindahan yang lebih baik daripada kecerdasan.”

(Nabi Muhammad)

PERSEMBAHAN

Untuk kedua orangtuaku Bapak Amad Daruji dan Ibu Siti Aminah

Untuk kakakku Ayu dan adikku Amanda

Untuk teman-teman Matematika 2014

Untuk Universitas Negeri Semarang

PRAKATA

Puji syukur peneliti panjatkan atas kehadiran Tuhan Yang Maha Esa yang telah melimpahkan kasih dan karunia-Nya, sehingga peneliti dapat menyelesaikan skripsi yang berjudul **“APLIKASI MODEL REGRESI PH COX UNTUK ANALISIS TAHAN HIDUP MENGGUNAKAN PENDEKATAN BAYESIAN”**.

Dalam penyusunan skripsi ini tidak lepas dari kendala dan hambatan, namun berkat bimbingan dan motivasi dari semua pihak yang telah membantu peneliti dalam menyelesaikan penyusunan skripsi ini, maka peneliti menyampaikan ucapan terima kasih khususnya kepada:

1. Prof. Dr. Fathurrahman, M.Hum. selaku Rektor Universitas Negeri Semarang.
2. Prof. Dr. Sudarmin, M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si. selaku Ketua Jurusan Matematika Universitas Negeri Semarang
4. Drs. Mashuri, M.Si., selaku Ketua Program Studi Jurusan Matematika Universitas Negeri Semarang.
5. Dr. Nurkaromah Dwidayati, M.Si., selaku Dosen Pembimbing I yang telah memberikan petunjuk dan pengarahan dalam menyusun skripsi.
6. Drs. Arief Agoestanto, M.Si. selaku Dosen Pembimbing II yang telah memberikan petunjuk dan pengarahan dalam menyusun skripsi.

7. Dr. Scolastika Mariani, M.Si., selaku Dosen Penguji yang telah memberikan penilaian dan saran dalam perbaikan skripsi ini.
8. Kedua orang tua serta adik dan kakak yang dengan kasih memberikan doa, motivasi, dan semangat
9. Teman-teman yang sudah membantu dan memberi motivasi demi kelancaran skripsi ini.
10. Sahabat–sahabat terima kasih atas canda, tawa dan perjuangan yang telah dilakukan bersama-sama dalam penyusunan skripsi ini.

Semarang, Mei 2019

Penulis

ABSTRAK

Ningtias, Yunita Dwi Ayu. (2019). *Aplikasi Model Regresi PH Cox untuk Analisis Tahan Hidup Menggunakan Pendekatan Bayesian*. Skripsi, Jurusan Matematika Fakultas Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama: Nur Karomah Dwidayati, M.Si, dan pembimbing pendamping: Drs. Arief Agoestanto, M.Si.

Kata kunci: Analisis *survival*, Regresi Cox *Proportional Hazard*, Bayesian, *Tuberculosis* (TB)

Analisis *survival* adalah teknik statistik yang digunakan untuk analisis data dimana variabelnya adalah waktu dari awal pengamatan hingga terjadinya suatu kejadian atau sering disebut waktu *survival*. Salah satu metode regresi *survival* yang sering digunakan adalah regresi cox *proportional hazard*. Tujuan dari penelitian ini adalah memperoleh estimasi parameter menggunakan pendekatan Bayesian, mengetahui faktor-faktor yang mempengaruhi ketahanan hidup pasien Penyakit Tuberculosis serta membentuk model regresi Cox *proportional hazard*nya.

Dalam penelitian ini menggunakan model cox *proportional hazard* menggunakan pendekatan Bayesian untuk menganalisis data *survival* pasien penderita TB di RSUD Tugurejo Semarang sebanyak 70 pasien. Hasil penelitian menunjukkan bahwa faktor yang berpengaruh secara signifikan terhadap waktu *survival* pasien TB yaitu variabel umur dengan model *proportional hazard*nya $h(t)=h_0(t)\exp(0,009898x_1)$. Koefisien x_1 sebesar 0,009898, artinya semakin bertambah umur sebesar satu satuan maka penambahan nilai $h(t)$ sebesar kelipatan $\exp(0,009898)$ satuan.

DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN TULISAN	ii
HALAMAN PENGESAHAN.....	iii
MOTTO DAN PERSEMBAHAN	iv
PRAKATA.....	v
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	7
1.3 Tujuan Masalah	7
1.4 Manfaat Penelitian	7
1.5 Batasan Masalah	8
BAB II LANDASAN TEORI	
2.1 Analisis <i>Survival</i>	9
2.2 Penyensoran Data	10
2.3 Fungsi <i>Survival</i>	12
2.4 Fungsi <i>Hazard</i>	13

2.5	Distribusi Lognormal.....	13
2.6	Regresi Cox <i>Proportional Hazard</i>	15
2.7	Model Cox <i>Proporsional Hazard</i>	15
2.8	Asumsi <i>Proporsional Hazard</i>	16
2.9	Metode Bayesian	17
2.9.1	Distribusi <i>Prior</i>	18
2.9.2	Fungsi Likelihood.....	29
2.9.3	Distribusi <i>Posterior</i>	19
2.10	MCMC.....	21
2.11	<i>Gibbs Sampling</i>	22
2.12	WinBUGS.....	23
2.13	Tuberkulosis (TB).....	24
2.14	Kerangka Berpikir	26
 BAB III METODE PENELITIAN		
3.1	Populasi dan Sampel Penelitian.....	32
3.2	Variabel Penelitian.....	32
3.3	Pengumpulan Data.....	34
3.4	Metode Pengolahan Data.....	35
 BAB IV HASIL & PEMBAHASAN		
4.1	Karakteristik Pasien <i>Tuberkulosis</i> (TB)	39
4.2	Pengecekan asumsi <i>Proportional Hazard</i> (PH)	43
4.3	Distribusi Data.....	44
4.4	Estimasi Parameter Distribusi Penduga.....	46

4.5	Estimasi Parameter Model Regresi Proportional Hazard	49
4.6	Analisis Data.....	52
4.6.1	Estimasi Parameter Distribusi Lognormal	52
4.6.2	Fungsi Survival dan Fungsi Hazard	54
4.6.3	Pembentukan Model Regresi	55
4.7	Aplikasi	58
4.8	Pembahasan	59
BAB V PENUTUP		
5.1	Kesimpulan.....	63
5.2	Saran	63
DAFTAR PUSTAKA		65
LAMPIRAN.....		71

DAFTAR TABEL

Tabel 3.1 Variabel Penelitian.....	34
Tabel 4.1 Persentase data pasien TB di RSUD Tugurejo Semarang	39
Tabel 4.2 Deskriptif Waktu Uji Hidup Pasien <i>Tuberkulosis (TB)</i>	39
Tabel 4.3 Deskriptif Pasien <i>Tuberkulosis (TB)</i>	40
Tabel 4.5 Nilai <i>p-value Global Test</i>	43
Tabel 4.4 <i>Anderson Darling</i> dan <i>P-value</i> Distribusi.....	45
Tabel 4.6 Estimasi parameter distribusi lognormal.....	54
Tabel 4.7 Estimasi Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i>	55
Tabel 4.8 Estimasi Parameter Model Regresi	57

DAFTAR GAMBAR

Gambar 4.1 pie chart variabel Jenis Kelamin pasien TB	40
Gambar 4.2 pie chart variabel Status Gizi pasien TB	41
Gambar 4.3 pie chart variabel Riwayat TB pasien TB	41
Gambar 4.4 pie chart variabel Batuk Darah pasien TB	42
Gambar 4.5 pie chart variabel Penyakit Penyerta pasien TB	42
Gambar 4.6 <i>Scatter Plot</i> data Pasien TB	44
Gambar 4.7 <i>Diagnostic plot</i> parameter distribusi lognormal	53
Gambar 4.8 <i>Diagnostic plot</i> parameter regresi	56

DAFTAR LAMPIRAN

Lampiran 1	Data Penelitian Waktu Pasien <i>Tuberculosis</i> (TB) yang Berobat Di RSUD Tugurejo Semarang Tahun 2017	71
Lampiran 2	Data Penelitian Pasien <i>Tuberculosis</i> (TB) di RSUD Tugurejo Semarang tahun 2017	73
Lampiran 3	Output Deskriptif Pasien TB	76
Lampiran 4	Output Pie Chart Pasien TB	77
Lampiran 5	Output Mempersiapkan Data (<i>setting time and event</i>).....	78
Lampiran 6	Output Global Test.....	79
Lampiran 6	Output distribusi waktu survival.....	85
Lampiran 8	Code WinBUGS Estimasi Parameter Distribusi Lognormal	86
Lampiran 9	<i>Output</i> WinBUGS Estimasi Parameter Distribusi Lognormal	87
Lampiran 10	Output SPSS Generalized Linear Model	89
Lampiran 11	Code WinBUGS Estimasi Parameter Regresi	90
Lampiran 12	Output WinBUGS Estimasi Parameter Regresi	92
Lampiran 13	Output Baseline Survival dan Baseline Hazard	97
Lampiran 14	Surat ijin Observasi dari RSUD Tugurejo Semarang	99s

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika merupakan suatu alat yang memegang peranan penting untuk mengambil suatu keputusan. Ilmu statistika sering sekali digunakan dalam berbagai bidang misalnya bidang ekonomi, industri, pendidikan, dan kesehatan. Salah satu contoh teknik statistika dalam bidang kesehatan yaitu analisis data uji hidup yang biasa disebut Analisis *Survival*. Analisis *survival* digunakan untuk mengetahui lamanya tahan hidup seorang pasien dalam pengobatan suatu penyakit, untuk melihat seberapa efektif suatu obat atau metode pengobatan yang diterapkan pada pasiennya, serta untuk mengetahui faktor apa saja yang mempengaruhi tahan hidup pasien terhadap penyakit.

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*. Analisis *survival*/analisis data ketahanan hidup adalah suatu metode yang berhubungan dengan waktu, mulai dari awal pengamatan sampai terjadinya kejadian khusus. Analisis ini bertujuan untuk mempelajari dan memodelkan hubungan antara faktor risiko dan waktu terjadinya kejadian khusus.

Analisis *survival* memerlukan data yang merupakan waktu ketahanan dari suatu individu. Data bisa diperoleh dari suatu pengamatan terhadap individu yang diamati waktu terjadinya kegagalan. Kegagalan yang dimaksud antara lain adalah

kematian karena penyakit tertentu, kambuhnya suatu penyakit atau munculnya penyakit baru.

Model regresi *Cox* merupakan salah satu model yang dapat digunakan dalam analisis *survival* untuk data dimana asumsi *proportional hazard* pada semua peubah bebas terpenuhi. Model ini termasuk ke dalam regresi semiparametrik yang bertujuan mengetahui kombinasi faktor-faktor yang berpengaruh terhadap responnya yang berupa waktu *survival*. Model *Cox* sering digunakan daripada metode lainnya karena dapat mengestimasi *hazard* ratio tanpa perlu diketahui fungsi *hazard* dasarnya, serta model *Cox* hampir sama dengan hasil model parametrik.

Kelebihan dari regresi ini yaitu tidak memiliki asumsi mengenai sifat dan bentuk sesuai dengan distribusi normal seperti asumsi pada regresi yang lain, distribusi yang digunakan adalah sesuai dengan distribusi dari variabel dependen yang digunakan. Sehingga baik digunakan bila distribusi dari waktu *survival* tidak diketahui secara pasti sehingga hasil estimasi parameter regresi masih dapat dipercaya. Asumsi pemodelan hanya memvalidasi asumsi bahwa fungsi *hazard* harus proporsional setiap waktu.

Untuk mengestimasi parameter dalam model dikenal ada dua pendekatan yaitu pendekatan klasik dan pendekatan Bayesian. Pada penelitian ini akan digunakan pendekatan Bayesian. Keuntungan yang diperoleh dalam penggunaan metode Bayes adalah memberikan pendugaan yang lebih baik daripada pendugaan dalam metode klasik. Karena penaksiran parameter dengan metode klasik sepenuhnya berdasarkan informasi dari data sampel, dimana ukuran sampel sangat

berpengaruh terhadap hasil pendugaan. Sedangkan pendugaan parameter dalam metode Bayes, selain menggunakan informasi dari data sampel juga memperhitungkan suatu distribusi awal yang disebut distribusi prior untuk mendapatkan sebaran posterior, sehingga hasil penduga akan jauh lebih baik.

Keuntungan lain dari penggunaan pendekatan bayesian, yaitu pendekatan Bayesian mampu menyediakan inferensia dan estimasi yang stabil meski pada situasi dimana ukuran sampel sangat kecil (Hoff, 2009).

Dalam perkembangannya, estimasi parameter dengan pendekatan bayesian ini dapat dilakukan dengan cara simulasi *Markov Chain Monte Carlo (MCMC)*, khususnya teknik *Gibbs Sampling*. Dengan pendekatan ini kompleksitas dalam model lebih mudah untuk diatasi. Alat bantu perhitungan menggunakan paket program winBUGS 1.4 yang telah memuat algoritma *BUGS (Bayesian Interface Using Gibbs Sampling)*.

Subjek dalam penelitian ini adalah pasien penderita penyakit *tuberkulosis* yang dirawat di RSUD Tugurejo Semarang. *Tuberkulosis (TB)* adalah suatu penyakit menular yang disebabkan oleh kuman dari kelompok *Mycobacterium* yaitu *Mycobacterium Tuberculosis*. Sebagian besar kuman TB menyerang paru, tetapi dapat juga mengenai organ tubuh lainnya. TB sampai saat ini masih merupakan salah satu masalah kesehatan masyarakat di dunia walaupun upaya pengendalian dengan strategi *Directly Observed Treatment (DOTS)* telah diterapkan di banyak negara sejak tahun 1995 (Depkes,2014).

Menurut *World Health Organization (WHO)* penyakit *Tuberkulosis Paru (TB)* saat ini telah menjadi ancaman global, karena hampir sepertiga penduduk

dunia telah terinfeksi. Sebanyak 95% kasus TB dan 98% kematian akibat TB didunia, terjadi pada negara-negara berkembang. Berdasarkan data *World Health Organization* (WHO) pada tahun 2013 terdapat 9 juta penduduk dunia telah terinfeksi kuman TB. Kemudian pada tahun 2014 terdapat 9,6 penduduk dunia terinfeksi kuman TB.

Laporan WHO pada tahun 2013, mencatat peringkat Indonesia menurun ke posisi lima dengan jumlah penderita TBC sebesar 429 ribu orang. Lima negara dengan jumlah kasus terbesar pada tahun 2013 adalah India, Cina, Afrika Selatan, Nigeria dan Indonesia.

Jumlah kematian akibat tuberkulosis paru diperkirakan 61.000 kematian per tahunnya (WHO, 2011). Menurut laporan Depkes tahun 2013, jumlah kasus tertinggi penderita Tuberkulosis terdapat di provinsi dengan jumlah penduduk yang padat seperti Jawa Barat, Jawa Timur, dan Jawa Tengah.

Menurut dinas kesehatan Kota Semarang tahun 2016, Penderita TB BTA (semua tipe) pada tahun 2016 sejumlah 3.251 kasus, dengan persentase TB Semua Tipe pada laki-laki sebanyak 1.845 kasus (57 %) lebih besar dari pada perempuan sebanyak 1.406 kasus (43 %). Hal ini disebabkan karena (fakta kualitatif) pada laki-laki lebih intens kontak dengan faktor risiko dan kurang peduli terhadap aspek pemeliharaan kesehatan individu dibandingkan dengan wanita.

Berdasarkan angka temuan kasus Tuberkulosis Paru yang didapat dari rumah sakit Tugurejo Semarang dari mulai tahun 2008 sampai 2012 yaitu pada tahun 2008 terdapat 313 pasien, tahun 2009 (380 pasien), tahun 2010 (349

pasien), tahun 2011 (420 pasien), dan terakhir pada bulan Maret 2012 terdapat 127 pasien penderita TB. Dilihat dari data tersebut setiap tahunnya jumlah pasien TB telah mengalami peningkatan.

Dengan meningkatnya jumlah kasus TB di tiap tahunnya, maka dibutuhkan penelitian mengenai faktor-faktor yang mempengaruhi waktu *survival* penderita penyakit TB dengan menggunakan analisis survival.

Beberapa penelitian sebelumnya oleh Tuan Hanni dan Triastuti Wuryandari (2013), Model Cox Proporsional Hazard digunakan untuk menganalisis dan menentukan survival rate dari data tersensor yang dipengaruhi satu atau lebih variabel penjelas. Model ini mengasumsikan bahwa hazard rate kelompok proporsional dengan hazard kelompok lain. Penelitian berikutnya oleh Ni Putu Ernawatiningsih (2012), penelitian ini memodelkan waktu survival. Metode yang digunakan adalah Analisis *Survival* dengan model regresi cox.

Kedua penelitian tersebut menggunakan metode maximum likelihood estimation (MLE) untuk mengestimasi parameter. Metode ini sering digunakan karena teknik estimasi parameternya yang lebih mudah. Tetapi teknik ini tidak bisa digunakan bila distribusi populasinya tidak diketahui. Selain itu teknik ini sensitif terhadap data ekstrim. Data ekstrim ini sangat berpengaruh terhadap nilai rata-rata dan variansi. Karena kekurangan dari metode maximum likelihood, para peneliti mencoba menggunakan metode lain yang bisa mengatasi kekurangan dari metode maximum likelihood yakni metode Bayesian. Menurut Bayes, parameter populasi berasal dari suatu distribusi sehingga nilainya tidaklah tunggal (merupakan variabel random). Karena nilai parameternya berasal dari suatu

distribusi, maka kesulitan pertama yang dijumpai adalah bagaimana bentuk distribusi parameter tersebut. Walaupun untuk menentukan distribusi prior dari parameter itu sulit, tetapi estimasi parameter dengan metode Bayes tampaknya lebih menjanjikan karena peneliti tidak perlu tahu tentang distribusi prior dari populasi.

Penelitian menggunakan metode Bayes diantaranya Sri Astuti Thamrin dkk (2018), dilakukan perbandingan nilai estimasi menggunakan metode MLE dan metode Bayesian pada data simulasi survival berdistribusi Weibull. Berdasarkan nilai MSE yang diperoleh menunjukkan bahwa metode Bayesian merupakan penaksir yang lebih baik. Hal ini karena metode Bayesian memiliki nilai MSE yang lebih kecil dibandingkan dengan metode MLE dengan menggunakan beberapa ukuran sampel yang berbeda-beda. Dwi Nurlaila dkk (2013), penelitian ini membandingkan metode MLE dan metode Bayes dalam menduga parameter Distribusi Eksponensial. Perbandingan kedua metode dilakukan melalui simulasi data pada berbagai kondisi parameter dan ukuran sampel. Evaluasi terhadap kedua metode dilakukan melalui pengamatan terhadap nilai bias dan MSE yang dihasilkan. Penelitian ini menunjukkan bahwa metode Bayes dengan nilai konstanta Jeffrey lebih kecil dari satu selalu menghasilkan nilai bias dan MSE yang lebih baik dibandingkan dengan metode Maximum Likelihood Estimation (MLE).

Berdasarkan penelitian tersebut, dalam penelitian ini menggunakan model cox proportional hazard dengan pendekatan Bayesian untuk menganalisis data *survival* pasien penderita TB.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah diuraikan pada bagian latar belakang, maka permasalahan yang timbul antara lain.

1. Bagaimana memperoleh estimasi parameter model regresi *Cox proporsional hazard* menggunakan metode Bayesian?
2. Apa saja faktor-faktor yang mempengaruhi ketahanan hidup pasien Penyakit Tuberkulosis?
3. Bagaimana memperoleh model regresi *Cox proporsional hazard* pasien Penyakit Tuberkulosis?

1.3 Tujuan Penelitian

Berdasarkan permasalahan di atas, maka tujuan penelitian ini adalah sebagai berikut.

1. Memperoleh estimasi parameter model regresi *Cox proporsional hazard* menggunakan metode Bayesian.
2. Mengetahui faktor-faktor yang mempengaruhi ketahanan hidup pasien Penyakit Tuberkulosis.
3. Memperoleh model regresi *Cox proporsional hazard* pasien Penyakit Tuberkulosis.

1.4 Manfaat Penelitian

Manfaat yang bisa diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi Penulis

Menambah ilmu tentang analisis *survival* dengan pendekatan Bayesian dan penerapannya, dan dapat menerapkan ilmu pengetahuan dan

metodologi yang telah diperoleh selama di bangku kuliah dalam penerapan di dunia nyata.

2. Bagi Pembaca

Hasil penelitian ini diharapkan dapat menambah pengetahuan tentang. Pengaplikasian Model Regresi *Cox* untuk Analisis Tahan Hidup dengan pendekatan Bayesian. Dan diharapkan dapat berguna bagi peneliti lainnya untuk melakukan penelitian lebih lanjut.

3. Bagi Rumah Sakit

Sebagai bahan pertimbangan dan masukan bagi RSUD Tugurejo Semarang dalam rangka meningkatkan pelayanan kesehatan khususnya penderita Penyakit Tuberkulosis. Serta dapat dijadikan sebagai salah satu sumber informasi yang dapat mendukung tujuan bagi Rumah Sakit.

1.5 Batasan Penelitian

Data yang digunakan yaitu waktu ketahanan hidup atau lama rawat inap pasien Penyakit Tuberkulosis yang dirawat di RSUD Tugurejo pada Januari sampai Desember 2017 dan faktor-faktor yang diduga mempengaruhi yaitu Umur, jenis kelamin, status gizi, riwayat TB, batuk darah, dan penyakit penyerta. Permasalah dibatasi pada pembentukan model regresi *Cox proportional hazard* dengan faktor-faktor yang mempengaruhi ketahanan hidup pasien penderita Penyakit Tuberkulosis. Penaksiran parameternya menggunakan metode Bayesian.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Survival

Analisis survival adalah teknik statistik yang digunakan untuk analisis data dimana variabelnya adalah waktu dari suatu objek yang dimulai dari awal pengamatan hingga terjadinya suatu peristiwa (*event*) atau sering disebut waktu survival (Kleinbaum, 2005). Peristiwa itu dapat berupa kematian, kesembuhan, kekambuhan dan lain-lain (Lee, 1992).

Namun yang sering terjadi, suatu individu tidak mengalami suatu peristiwa sampai penelitian berakhir. Hal tersebut yang mengakibatkan ketidaklengkapan data kegagalan (*failure time*) yang sering disebut *censoring* (*censoring*) (Jaya, 2015).

Analisis *survival* bertujuan untuk memodelkan distribusi dari variabel waktu kegagalan (*failure time*) dan menaksirkan ketergantungan variabel waktu kegagalan dengan variabel bebas (Thamrin, 2008).

Menurut Kleinbaum (2005) Meskipun lebih dari satu peristiwa bisa dianggap dalam analisis yang sama, disini akan diasumsikan bahwa hanya satu peristiwa yang diminati. Bila terjadi lebih dari satu peristiwa, maka akan dipertimbangan (misalnya, kematian karena penyebab lain). Dalam analisis survival, biasanya mengacu pada variabel waktu sebagai waktu survival, karena ini menunjukkan waktu bahwa seseorang telah “survived” selama beberapa periode tindak lanjut. Kita juga biasanya mengacu pada kejadian tertentu sebagai

kegagalan, karena kejadian penting adalah kematian, kejadian penyakit, atau beberapa hal negatif lainnya kejadian seseorang.

Ada tiga faktor yang harus diperhatikan dalam menentukan waktu *survival* menurut Cox (1972):

- (1) *Time origin or starting point* (titik awal) adalah waktu ketika suatu objek masuk dalam penelitian. Titik awal ini misalnya awal mula pengamatan berupa waktu dimulainya suatu pengobatan baru atau tanggal masuk rawat inap Rumah Sakit.
- (2) *Ending event of interest* (kejadian akhir) adalah kejadian yang menjadi inti dari penelitian. Kejadian akhir didefinisikan jelas untuk mengetahui status tersensor atau tidak tersensor, meninggal atau sembuh seorang pasien.
- (3) *Measurement scale for the passage of time* (skala pengukuran). Skala pengukuran sebagai batas dari waktu kejadian dari awal sampai akhir kejadian, misalnya skala tahunan, bulanan, mingguan, atau harian.

Analisis *survival* memiliki tipe data lama waktu yang tersensor sehingga tidak dapat menggunakan analisis regresi linier biasa dan tidak pula dapat menggunakan analisis regresi logistik karena lamanya waktu data yang diikuti tidak sama sehingga yang paling sesuai adalah menggunakan analisis *survival* (Vittinghoff, et al., 2005 dalam Rejki, 2015).

2.2 Penyensoran Data

Sumber kesulitan pada analisis *survival* yaitu adanya data tersensor, yaitu adanya kemungkinan beberapa individu tidak dapat diobservasi secara penuh sampai terjadinya *failure event*, sehingga beberapa data observasi tidak dapat

diperoleh informasinya secara lengkap (Le, 1997). Jadi, data dikatakan tersensor jika pengamatan waktu *survival* hanya diperoleh sebagian informasi atau kejadian akhirnya tidak seluruhnya dapat diamati (Collet, 2003).

Menurut Kleinbaum (2005) survival analisis harus mempertimbangkan masalah analitik utama yang disebut penyensoran. Adapun tujuan penyensoran yaitu memperpendek waktu percobaan karena dalam mengukur waktu kegagalan atau kematian individu, kadang diperlukan waktu yang lama dan biaya yang tidak sedikit.

Menurut Collet (2003), terdapat tiga penyebab terjadinya data tersensor, antara lain :

- (1) *Loss to follow up*, bila objek pindah, meninggal, atau menolak untuk berpartisipasi.
- (2) *Drop out*, bila perlakuan dihentikan karena suatu alasan tertentu.
- (3) *Termination of study*, bila masa penelitian berakhir sementara objek yang diobservasi belum mencapai *failure event*.

Ada tiga jenis penyensoran menurut Lee & Wang (2003):

(1) Penyensoran Tipe I

Sensor tipe I adalah tipe penyensoran dimana percobaan akan dihentikan setelah mencapai waktu T yang telah ditentukan untuk mengakhiri semua n individu yang masuk pada waktu yang sama. Berakhirnya waktu uji T menjelaskan waktu sensor uji. Dengan kata lain, jika tidak terdapat individu yang hilang secara tiba-tiba maka waktu tahan hidup observasi tersensor sama dengan lama waktu pengamatan.

(2) Penyensoran Tipe II

Sensor tipe II adalah tipe penyensoran dimana percobaan dihentikan setelah diperoleh sejumlah r kegagalan tertentu dari n unit ($1 \leq r \leq n$). Pada data tersensor tipe II, individu masuk ke dalam penelitian pada waktu yang sama dan penelitian dihentikan jika sejumlah individu yang telah ditentukan gagal, mati, atau rusak.

(3) Penyensoran Tipe III

Pada sensor tipe III, individu atau unit uji masuk ke dalam percobaan pada waktu yang berlainan selama periode waktu tertentu. Beberapa unit uji mungkin gagal atau mati sebelum pengamatan berakhir sehingga waktu tahan hidupnya dapat diketahui secara pasti. Kemungkinan kedua adalah unit uji keluar sebelum pengamatan berakhir, atau kemungkinan ketiga adalah unit uji tetap hidup sampai batas waktu terakhir pengamatan.

2.3 Fungsi *Survival*

Jika T merupakan variabel random dari waktu *survival* suatu individu dalam interval $[0, \infty)$, maka fungsi distribusi kumulatif $F(t)$ untuk distribusi kontinu dengan fungsi kepadatan peluang $f(t)$ dinyatakan sebagai berikut:

$$F(t) = P(T \leq t) = \int_0^t f(t) dt, \text{ untuk } t > 0 \quad (2.1)$$

Fungsi *survival* ($S(t)$) merupakan probabilitas suatu individu untuk bertahan hidup sampai dengan waktu t dengan ($t \geq 0$) atau fungsi *survival* adalah peluang waktu *survival* lebih besar dari t . Fungsi *survival* merupakan komplemen dari variabel random fungsi distribusi kumulatif $F(t)$. Oleh karena itu diperoleh fungsi ketahanan hidup (*survival*) yang didefinisikan dengan

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \quad (2.2)$$

2.4 Fungsi Hazard

Fungsi hazard ($h(t)$) adalah peluang suatu individu meninggal atau mengalami kejadian sesaat dalam interval waktu t sampai $t+\Delta t$, jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t .

Fungsi hazard secara matematika dinyatakan sebagai:

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t} \right\} \quad (2.3)$$

Fungsi hazard juga dapat dinyatakan dengan menggunakan fungsi distribusi $F(t)$ dan fungsi kepadatan peluang $f(t)$, yaitu

$$h(t) = \frac{f(t)}{1-F(t)}, \text{ dengan syarat } F(t) \neq 1 \quad (2.4)$$

Dari persamaan (2.3) dan (2.5) diperoleh

$$h(t) = \frac{f(t)}{S(t)} \quad (2.5)$$

Selanjutnya untuk fungsi hazard kumulatif didefinisikan dengan peluang hazard dari interval 0 sampai t yang dinyatakan dengan

$$H(t) = \int_0^t h(x) dx \quad (2.6)$$

2.5 Distribusi Lognormal

Fungsi kepadatan peluang dari distribusi lognormal adalah :

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\ln t - \mu}{\sigma} \right)^2 \right], \sigma > 0, t > 0 \quad (2.7)$$

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

$$F(t) = \int_{-\infty}^t f(t) dt = \int_{-\infty}^t \frac{1}{\sigma t \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\ln t - \mu}{\sigma} \right)^2 \right] dt$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t \frac{1}{t} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right] dt$$

dimisalkan, $\frac{\ln t - \mu}{\sigma} = u$

$$\ln t = u\sigma + \mu$$

$$\frac{1}{t} dt = \sigma du$$

maka,

$$\begin{aligned} F(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t \exp\left[-\frac{1}{2}u^2\right] \sigma du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left[-\frac{1}{2}u^2\right] du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \exp\left[-\frac{1}{2}u^2\right] du + \frac{1}{\sqrt{2\pi}} \int_0^t \exp\left[-\frac{1}{2}u^2\right] du \\ &= \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln t - \mu}{\sigma\sqrt{2}}\right) \end{aligned} \quad (2.8)$$

Nilai $F(t)$ pada persamaan (2.8) terdapat hasil erf dimana erf merupakan *error function* dari $F(t)$, hal ini dikarenakan $\frac{1}{\sqrt{2\pi}} \int_0^t \exp\left[-\frac{1}{2}u^2\right] du$ memiliki hasil yang tidak *close form* sehingga terbentuklah *error function*.

Kemudian untuk fungsi survivalnya diperoleh sebagai berikut :

$$S(t) = 1 - F(t) = 1 - \left(\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln t - \mu}{\sigma\sqrt{2}}\right)\right) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\ln t - \mu}{\sigma\sqrt{2}}\right) \quad (2.9)$$

Dan fungsi hazardnya adalah :

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{\sigma t \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right]}{\frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\ln t - \mu}{\sigma\sqrt{2}}\right)} = \frac{\sqrt{2} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right]}{\sigma\sqrt{\pi} \left(1 - \operatorname{erf}\left(\frac{\ln t - \mu}{\sigma\sqrt{2}}\right)\right)} \quad (2.10)$$

2.6 Regresi Cox Proportional Hazard

Regresi Cox Proportional Hazard digunakan untuk mengetahui efek dari beberapa variabel independent terhadap variabel dependent. Variabel dependent dalam regresi Cox adalah waktu survival suatu objek terhadap suatu peristiwa tertentu. 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 (Cox, 1972).

Regresi ini tidak memiliki asumsi mengenai sifat dan bentuk sesuai dengan distribusi normal seperti asumsi pada regresi yang lain, distribusi yang digunakan adalah sesuai dengan variabel dependen yang digunakan. Sehingga baik digunakan bila distribusi dari waktu survival tidak diketahui secara pasti sehingga hasil estimasi parameter regresi masih dapat dipercaya (Lee & Wang, 2003).

Menurut Dahlan (2012) asumsi yang harus dipenuhi dalam menggunakan regresi cox adalah dengan pengecekan asumsi *proportional hazard* (PH), yaitu variabel prediktor independen terhadap waktu dan hubungan antara kumulatif hazard konstan pada setiap waktu.

2.7 Model Cox Proporsional Hazard

Model *Cox proportional hazards* merupakan model yang umum digunakan untuk analisis *survival*. Model ini pertama kali dikenalkan oleh Cox. Model Cox Proporsional Hazard biasanya ditulis dalam bentuk formula model hazard berikut (menurut Kleinbaum, 2005):

$$h(t) = h_0(t)\exp(\beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p)$$

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right) \quad (2.11)$$

dimana

$h_0(t)$: nilai *baseline hazard* yang merupakan nilai fungsi hazard pada saat semua variabel bebas bernilai 0. Fungsi ini hanya melibatkan waktu survival t

X_1, X_2, \dots, X_p : nilai dari variabel bebas

$\beta_1, \beta_2, \dots, \beta_p$: parameter regresi

p : banyaknya variabel bebas

Formula model Cox memiliki sifat bahwa jika semua X sama dengan nol, rumusnya menjadi fungsi baseline hazard.

$$X_1 = X_2 = \dots = X_p = 0$$

$$\begin{aligned} h(t, \mathbf{X}) &= h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right) \\ &= h_0(t) \exp(0) \\ &= h_0(t) \text{ Baseline hazard} \end{aligned}$$

Tidak ada X dalam model: $h(t, \mathbf{X}) = h_0(t)$

Artinya, bagian eksponensial dari formula menjadi $\exp(0)$, yaitu 1. Sifat model Cox ini yang menjadi alasan mengapa $h_0(t)$ disebut fungsi baseline.

2.8 Asumsi Proporsional Hazard

Asumsi pemodelan yang harus dipenuhi dalam regresi *cox* adalah asumsi Proportional Hazard (PH). Dikatakan memenuhi asumsi PH apabila fungsi hazard konstan sepanjang waktu, artinya resiko suatu individu proportional terhadap individu lainnya dan independen terhadap waktu (Kleinbaum, 2005). Asumsi PH

sangat penting dalam analisis survival. Pentingnya analisis ini analog dengan asumsi normalitas data pada analisis parametrik (Dahlan, 2009).

Ada tiga jenis pengecekan asumsi Proportional Hazard, yaitu :

1. Garis survival pada kurva Kaplan-Meier tidak saling berpotongan
2. Garis survival pada $\ln-\ln$ survival tidak saling berpotongan
3. Uji *global test*

Pengecekan asumsi PH dengan global test pada data dilakukan sebagai berikut :

Hipotesis dari uji *global test* :

H_0 : Data memenuhi asumsi *PH*

H_1 : Data tidak memenuhi asumsi *PH*

Signifikansi α : 0,05

Penolakan H_0 jika $p - value < \alpha$ artinya data memenuhi asumsi *PH*

2.9 Metode Bayesian

Secara umum, estimasi yang digunakan dalam statistika inferensia didasarkan pada data sampel dalam populasi. Berbeda halnya dengan estimasi pada pendekatan Bayesian, dimana informasi dari data sampel dan distribusi awal yang disebut sebagai distribusi *prior* dikombinasikan untuk menghasilkan distribusi *posterior*, sehingga diperoleh estimator Bayesian yang merupakan mean dan modus dari distribusi *posterior*. Hasil yang dinyatakan dalam bentuk distribusi posterior yang kemudian menjadi dasar dalam metode bayesian (Ntzoufras, 2009).

Metode Bayesian dalam statistik memiliki perbedaan yang mendasar dengan metode klasik. Dalam metode klasik, parameter θ dipandang sebagai besaran sementara pada Bayesian parameter θ dianggap sebagai variabel random yang memiliki distribusi yang disebut dengan distribusi *prior* (Lee, 2007).

Berikut adalah penjelasan mengenai distribusi *prior*, fungsi *likelihood*, dan distribusi *posterior*.

2.9.1 Distribusi Prior

Di dalam analisis Bayesian, ketika suatu populasi mengikuti distribusi tertentu dengan suatu parameter didalamnya, misal θ , maka dimungkinkan bahwa parameter θ itu sendiri juga mengikuti suatu distribusi probabilitas tertentu, yang disebut sebagai distribusi *prior* (Bain dan Engelhardt, 1992).

Distribusi *prior* adalah distribusi awal yang memberikan informasi mengenai parameter. Distribusi *prior* mencerminkan kepercayaan subyektif parameter sebelum sampel diambil (Wulandari, 2015).

Distribusi *posterior* akan bergantung pada pemilihan distribusi *prior*. Dengan demikian permasalahan utama adalah bagaimana memilih distribusi *prior* untuk parameter yang tidak diketahui tetapi memiliki relevansi dengan permasalahan data penelitian. Dalam menentukan distribusi *prior* dapat dilihat berdasarkan ruang parameternya (Lukitasari, 2015).

Box dan Tiao (1973) membagi prior menjadi 2 kelompok berdasarkan fungsi *Likelihood*nya:

- a. Berkaitan dengan pola model *likelihood* dari datanya.

- 1) *Prior* konjugat, mengacu pada acuan analisis model terutama dalam pembentukan fungsi *likelihood* nya.
 - 2) *Prior* non-konjugat, pemberian *prior* pada model tidak mempertimbangkan pola pembentuk fungsi *likelihood*nya.
- b. Berkaitan dengan penentuan masing-masing parameter pada pola distribusi *prior* tersebut.
- 1) *Prior* informatif, jika terdapat informasi mengenai parameter model yang ada dan pemilihan *prior* ini didasarkan pada informasi dari data yang diperoleh. Pemberian nilai parameter pada distribusi *prior* ini akan sangat mempengaruhi bentuk distribusi *posterior*.
 - 2) *Prior* non-informatif, jika tidak mengandung informasi mengenai parameter dan pemilihan *prior* tidak didasarkan pada data yang ada.

2.9.2 Fungsi Likelihood

Menurut Bain & Engelhardt (1992) Fungsi *likelihood* adalah fungsi kepadatan probabilitas bersama dari n variabel random X_1, X_2, \dots, X_n dan dinyatakan dalam bentuk $f(x_1, x_2, \dots, x_n | \theta)$. Jika x_1, x_2, \dots, x_n tetap, maka fungsi *likelihood* adalah fungsi dari parameter θ dan dinotasikan dengan $L(\theta)$. Jika x_1, x_2, \dots, x_n menyatakan suatu sampel random dari $f(x|\theta)$, maka:

$$L(\theta) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) \quad (2.12)$$

2.9.3 Distribusi Posterior

Dalam estimasi Bayesian, setelah informasi tentang sampel diperoleh dan *prior* dapat ditentukan maka distribusi *posterior* nya dapat dicari dengan cara

mengalikan *prior* nya dengan informasi sampel yang diperoleh dari *likelihood* nya (Bolstad, 2007).

Box dan Tiao (1973) mendefinisikan bahwa distribusi *posterior* $f(\theta|x)$ adalah fungsi densitas bersyarat parameter θ jika diketahui nilai observasi x , dapat dituliskan sebagai :

$$f(\theta|x) = \frac{f(x, \theta)}{f(x)} \quad (2.13)$$

dimana $f(x, \theta)$ adalah fungsi kepadatan bersama dari x dan θ sedangkan $p(x)$ merupakan distribusi marginal x .

Fungsi kepadatan bersama $f(x, \theta)$ merupakan perkalian dua dentitas, yaitu distribusi prior ($f(\theta)$) dan distribusi data ($f(x|\theta)$), yang ditulis sebagai :

$$f(x, \theta) = f(x|\theta) f(\theta) \quad (2.14)$$

Sedangkan distribusi marginal x dapat dihitung dengan :

$$f(x) = \begin{cases} \int f(x, \theta) d\theta = \int f(x|\theta) f(\theta) d\theta, & \theta \text{ kontinu} \\ \sum_{\theta} f(x, \theta) d\theta = \sum_{\theta} f(x|\theta) p(\theta), & \theta \text{ diskrit} \end{cases}$$

Sehingga persamaan 2.13 dapat dinyatakan sebagai

$$f(\theta|x) = \frac{f(x|\theta) f(\theta)}{f(x)} \quad (2.15)$$

Distribusi parameter θ yaitu $f(\theta)$ disebut sebagai prior, dan $f(x|\theta)$ sebagai distribusi sampling atau *likelihood* yang merupakan fungsi parameter dari θ .

Karena $f(x)$ tidak tergantung pada θ , maka $\frac{1}{f(x)}$ dapat dianggap konstan, misalkan

C . Dengan kata lain, $f(x)$ adalah suatu konstanta yang disebut sebagai *normalized constant*. Sehingga posterior dapat dituliskan sebagai berikut:

$$p(\theta|x) \propto l(x|\theta)p(\theta) \quad (2.16)$$

Persamaan (2.16) menunjukkan bahwa *posterior* adalah *proportional* terhadap *likelihood* dikalikan dengan *prior* dari parameter model.

Nilai tengah dari distribusi *posterior* yang akan digunakan untuk menentukan *estimator* dari parameter yang tidak diketahui (Soejoeti & Soebanar, 1988). Kesulitan juga dapat terjadi karena kompleksitas saat penentuan distribusi *posterior*, distribusi *posterior* susah untuk diestimasi secara langsung. Salah satu penyelesaian untuk mengatasi kesulitan tersebut adalah dengan pendekatan menggunakan metode simulasi MCMC (*Markov Chain Monte Carlo*) algoritma *Gibbs Sampling* (Lukitasari, 2015).

2.10 MCMC

Markov Chain Monte Carlo (MCMC) adalah suatu teknik metode simulasi yang membangkitkan sejumlah sampel dari distribusi data tertentu untuk mendapatkan distribusi *posterior* (Ntzoufras, 2009). Pada analisis Bayesian, penggunaan MCMC dapat mempermudah analisisnya, sehingga keputusan yang diambil dari hasil analisis akan dapat dilakukan dengan cepat dan tepat. Ada dua kemudahan yang diperoleh dari penggunaan metode MCMC pada analisis Bayesian. Pertama, metode MCMC dapat menyederhanakan bentuk integral yang kompleks dengan dimensi besar menjadi bentuk integral yang sederhana dengan satu dimensi. Kedua, dengan menggunakan metode MCMC, estimasi densitas data dapat diketahui dengan cara membangkitkan suatu rantai *Markov* yang berurutan sebanyak N yang cukup besar (Iriawan, 2001).

MCMC pada dasarnya adalah Integrasi Monte Carlo menggunakan Markov Chain. Metode ini merupakan metode simulasi untuk penarikan sampel dari distribusi posterior. Metode ini menarik sampel berturut-turut dari suatu distribusi targetnya. MCMC membangkitkan data sampel parameter θ yang memiliki distribusi tertentu melalui sebuah algoritma, dimana nilai setiap langkah bergantung pada satu langkah sebelumnya sehingga membentuk suatu rantai *Markov* (Gilks, 1996).

Terdapat dua metode MCMC yang paling populer yaitu algoritma Metropolis-Hastings (Metropolis *et al.*, 1953; Hastings, 1970) dan Gibbs sampling (Geman & Geman, 1984). Metode *Gibbs Sampling* merupakan teknik yang sering dipakai oleh pengguna metode Bayes (Walsh, 2004)

2.11 Gibbs Sampling

Salah satu pendekatan MCMC adalah dengan metode *Gibbs Sampling*. *Gibbs Sampling* diperkenalkan oleh Geman dan Geman (1984). *Gibbs Sampling* merupakan teknik simulasi untuk membangkitkan variabel acak dari distribusi tertentu secara langsung tanpa harus menghitung densitasnya. Dengan menggunakan *Gibbs Sampling*, penghitungan yang sulit dapat dihindari (Casella, *et al.*, 1992).

Gibbs sampling adalah algoritma MCMC yang mencakup iterative sampling dari tiap distribusi bersyarat, dimana parameter β dipartisi menjadi beberapa bagian, $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ dan distribusi posterior full conditional adalah $p(\beta_1|x, \beta_2, \dots, \beta_p), \dots, p(\beta_p|x, \beta_1, \dots, \beta_{p-1})$ atau secara sederhana dapat ditulis

$\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(p)}$ (Congdon, 2003). Gibbs sampling bekerja dengan langkah-langkah sebagai berikut.

a. Tentukan nilai awal dari $\beta^{(0)} = \{\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)}\}$

b. Bangkitkan

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

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

⋮

Nilai $\beta_p^{(1)}$ berasal dari $f(\beta_p | \beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_{p-1}^{(1)})$

Sehingga diperoleh parameter baru $(\beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_p^{(1)})$

c. Gunakan $(\beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_p^{(1)})$ sebagai nilai awal untuk proses selanjutnya sampai k iterasi yang mencapai konvergen.

d. Periksa konvergensi dengan melihat *trace plot*, *history plot*, *autocorrelation*, *quantile plot*, dan *density posterior*.

e. Mendapatkan mean, median, deviasi standar dari distribusi *posterior*.

2.12 WinBUGS

WinBUGS adalah perangkat lunak berbasis bahasa pemrograman yang digunakan untuk menghasilkan sampel acak dari distribusi *posterior* parameter model Bayes. Pengguna hanya perlu menentukan data, struktur dari model yang dipertimbangkan, dan beberapa nilai awal untuk parameter model. Tujuan utama dari proyek WinBUGS adalah untuk mengembangkan perangkat lunak untuk memproduksi sampel MCMC dari distribusi *posterior* parameter model yang diinginkan.

Nama winBUGS diambil dari isi paket programnya yang dikembangkan berdasarkan pada metode *Gibbs Sampling* dan dibuat untuk dapat dijalankan di dalam sistem operasi komputer *Windows*. Sehingga pengertian nama WinBUGS adalah *Bayesian Using Gibbs Sampling* (BUGS) dalam *Windows* (Ntzoufras, 2009). WinBUGS sekarang bisa dijalankan bahkan dari paket perangkat lunak lain, seperti R, Matlab, dan Excel.

2.13 Penyakit Tuberkulosis (TB)

Tuberkulosis (TB) adalah suatu penyakit infeksi saluran pernafasan menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Bakteri *mycobacterium tuberculosis* ini merupakan bakteri basil yang sangat kuat dan biasanya menyerang paru-paru, dan juga organ tubuh lainnya seperti kelenjar getah bening, usus, ginjal, kandung, tulang, bahkan otak (Rahajoe, 2005). Namun sebagian besar bakteri ini menyerang paru-paru dikarenakan penularan paling mudah melalui udara. Bakteri ini menyerang paru-paru sehingga pada bagian dalam alveolus yang menyebabkan proses difusi oksigen yang terganggu karena adanya bintik-bintik kecil di dinding alveolus. Apabila bagian paru-paru yang sudah terinfeksi meluas, maka sel-selnya mati dan paru-paru mengecil sehingga pernafasan penderita menjadi tidak stabil (Yurike, 2016).

Tingginya penularan dan infeksi TB paru berkaitan dengan beberapa faktor diterminan, diantaranya faktor lingkungan (tempat tinggal dan pekerjaan), karakteristik individu (umur, jenis kelamin, dan status gizi), perilaku (kebiasaan merokok), pengetahuan (pendidikan), dan adanya riwayat penyakit diabetes mellitus (DM) (Misnadiarly, 2006). Faktor lingkungan khususnya tempat tinggal

dan pekerjaan sangat berpengaruh terhadap angka kejadian TB paru. Bakteri penyebab TB paru dapat bertahan hidup dan tumbuh baik pada lingkungan kumuh atau kondisi rumah yang lembab dan kurang ventilasi serta akan mati pada suhu tinggi dengan paparan sinar ultraviolet. Adapun tempat kerja dengan tingkat polusi yang tinggi juga tidak bisa diabaikan. Apalagi pekerjaan yang mempunyai kontak langsung dengan mikobakterium TB seperti di rumah sakit, puskesmas, klinik maupun laboratorium (Soemantri, 2004).

Faktor lain yang mempengaruhi seseorang menjadi pasien TB adalah daya tahan tubuh rendah, diantaranya infeksi HIV/ AIDS dan malnutrisi (gizi buruk). Infeksi HIV dapat mengakibatkan kerusakan luas sistem daya tahan tubuh, sehingga jika terjadi infeksi penyerta (oportunistic), seperti tuberkulosis maka yang bersangkutan akan menjadi sakit parah bahkan bisa mengakibatkan kematian. Penyakit tuberkulosis harus mendapat penanganan yang tepat karena penyakit ini menyerang tidak memandang kelompok usia produktif, kelompok ekonomi lemah dan berpendidikan rendah. Penyakit TB paru lebih banyak ditemukan di daerah miskin. Karena faktor lingkungan yang kurang mendukung menjadi penyebab TB (Depkes RI,2010).

Sumber penularan adalah penderita TBC BTA Positif pada waktu batuk atau bersin, penderita menyebarkan kuman ke udara dalam bentuk droplet (percikan dahak). Droplet yang mengandung kuman dapat bertahan di udara pada suhu kamar selama beberapa jam dan mampu bertahan dalam keadaan dorman (tidak aktif) selama bertahun-tahun. Orang dapat terinfeksi jika droplet tersebut terhirup ke dalam saluran pernafasan. Setelah kuman TBC

masuk ke dalam tubuh manusia melalui pernafasan, kuman TBC tersebut dapat menyebar dari paru ke bagian tubuh lainnya (Depkes RI, 2002).

Obat tuberkulosis harus diminum oleh penderita secara rutin selama enam bulan berturut-turut tanpa henti. Kedisiplinan pasien dalam menjalankan pengobatan juga perlu diawasi oleh anggota keluarga terdekat yang tinggal serumah, yang setiap saat dapat mengingatkan penderita untuk minum obat. Apabila pengobatan terputus tidak sampai enam bulan, penderita sewaktu-waktu akan kambuh kembali penyakitnya dan kuman tuberkulosis menjadi resisten sehingga membutuhkan biaya besar untuk pengobatannya (Hiswani, 2006). Pengobatan TB bertujuan untuk menyembuhkan pasien, mencegah kematian, mencegah kekambuhan, memutuskan rantai penularan dan mencegah terjadinya resistensi kuman terhadap Obat Anti Tuberkulosis (OAT) (Depkes RI, 2011).

2.14 Kerangka Berpikir

Analisis *survival* (analisis data ketahanan hidup) merupakan suatu metode yang berhubungan dengan waktu, mulai dari awal pengamatan sampai terjadinya kejadian khusus, seperti kesembuhan atau kambuhnya penyakit atau bahkan kematian pasien. Metode-metode yang digunakan dalam analisis survival yaitu metode tabel hidup (*Life Table/Cutler-Ederer*), *Cox Proportional Hazard* atau Model *Cox* dan metode Product Limit (*Kaplan Meier*).

Menurut Gayatri (2005) Penggunaan metode *Life Table* yaitu dengan cara menentukan interval waktu yang dikehendaki. Syarat dan asumsi yang harus dipenuhi pada metode tersebut adalah saat awal pengamatan harus jelas, efek yang diteliti harus jelas dan harus berskala nominal dikotom atau dianggap sebanding

dengan pengukuran dengan skala kategorikal, kasus hilang masa pengamatan (*lost to follow*) harus independent terhadap efek, risiko untuk terjadi efek tidak bergantung pada tahun kalender, dan risiko untuk terjadi efek pada interval waktu yang dipilih dianggap sama. Sedangkan pada Kaplan Meier tidak dikelompokkan dalam interval tertentu sehingga dapat digunakan pada analisis *survival* dengan jumlah subjek yang sedikit dan efek yang dihitung tepat pada saat ia terjadi dengan anggapan bahwa keadaan tersebut sebanding dengan pengukuran berskala numerikal. Kelebihan dibanding *Life Table* adalah dapat memberikan proporsi ketahanan hidup yang pasti karena menggunakan waktu *survival* secara tepat bukan berdasarkan kelas interval.

Metode regresi Cox lebih terkenal dibandingkan dengan *Kaplan Meier* dan *Life Table* karena (Kleinbaum, 1996) dapat mengestimasi hazard ratio tanpa perlu diketahui $h_0(t)$, dapat mengestimasi $h_0(t)$, $h(t, X)$, dan fungsi *survival* meskipun $h_0(t)$ tidak spesifik, serta hasil dari Cox model hampir sama dengan hasil model parametrik. Penaksiran fungsi *hazard* dapat dipakai untuk menghitung risiko relatif terjadinya kejadian.

Metode *Life Table* dan *Kaplan Meier* hanya digunakan untuk mengestimasi dan menampilkan kurva waktu ketahanan saja. Kedua metode juga menampilkan median waktu ketahanan sebagai ukuran rerata waktu ketahanan. Sedangkan untuk regresi Cox, metode yang lebih terkenal daripada metode lainnya bisa digunakan untuk menghitung fungsi *hazard*. Karena dalam penelitian ini akan dicari ketahanan hidup individu dan faktor-faktornya maka akan lebih cocok jika menggunakan regresi Cox.

Terdapat banyak distribusi untuk menganalisis data survival, diantaranya ada eksponensial, weibull, gamma, log-normal, dan log-logistik. Akan dicari distribusi yang sesuai dengan data survival penderita penyakit TB dengan menggunakan *global test*.

Dalam teori estimasi, terdapat dua pendekatan yang sering digunakan yaitu pendekatan statistika klasik dan pendekatan Bayesian. Melalui pendekatan statistika klasik, penaksiran parameter sepenuhnya berdasarkan data sampel yang diambil dari populasi. Sedangkan melalui pendekatan Bayesian, disamping memanfaatkan data sampel diperoleh dari populasi juga memperhitungkan suatu distribusi awal yang disebut distribusi *prior* (Wong, et al, 2005).

Metode Bayesian dalam statistik memiliki perbedaan yang mendasar dengan metode klasik. Dalam metode klasik, parameter θ dipandang sebagai besaran yang bernilai tetap. Sementara pada Bayesian parameter θ dianggap sebagai variabel random yang memiliki distribusi yang disebut distribusi *prior* (Lee, 2007). Informasi dari sampel penelitian sebagai *likelihood* digunakan sebagai bahan untuk memperbaharui informasi *prior* menjadi sebuah *posterior* yang akan digunakan untuk bahan inferensi. Dengan kata lain metode Bayes merupakan sebuah metode yang optimal tentang bagaimana memperbaharui kepercayaan seseorang tentang parameter populasi setelah diberikannya informasi baru (Hoff, 2009). Dengan alasan tersebut, pada penelitian ini akan menggunakan metode Bayesian untuk mengestimasi parameter.

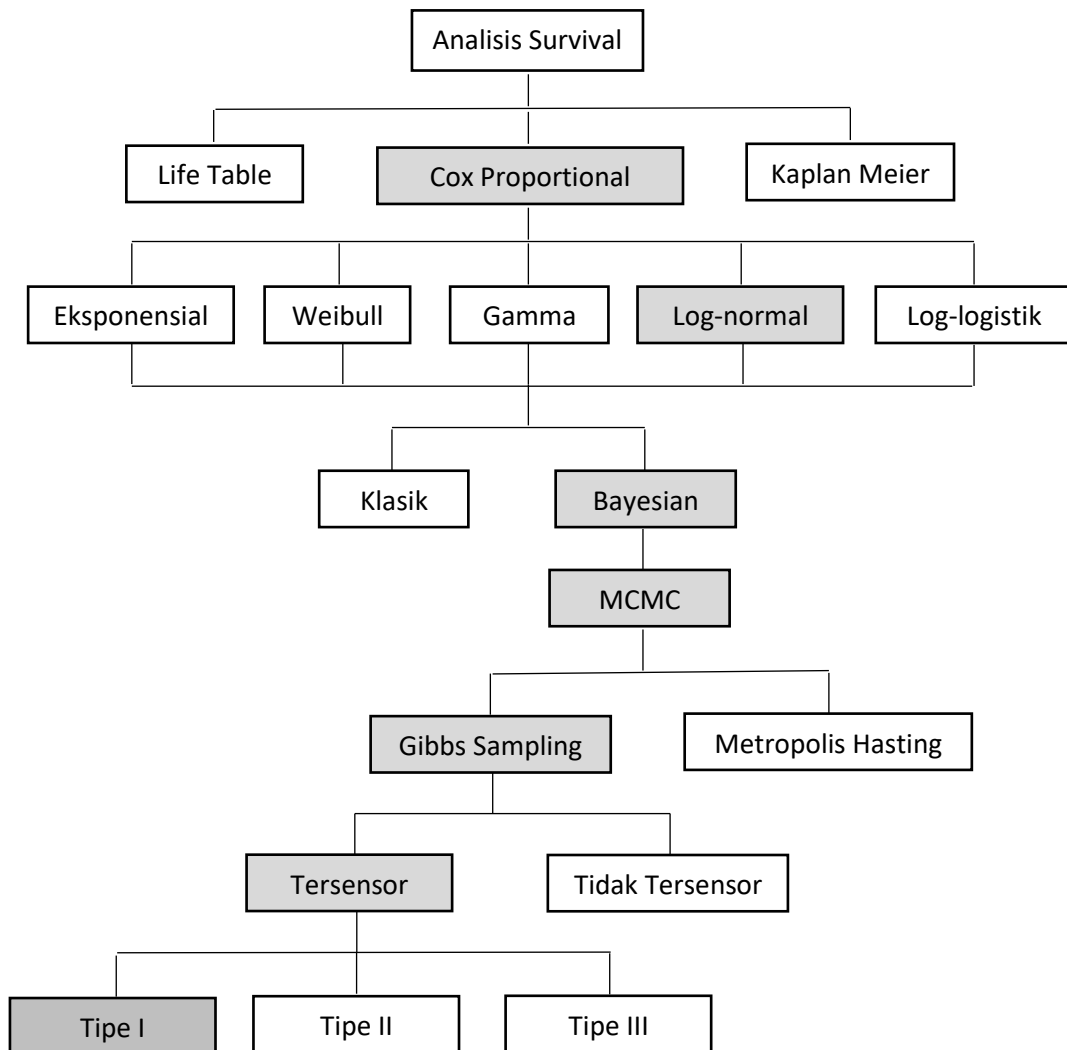
Pada penggunaan metode Bayesian, ada kalanya dihadapkan pada kondisi dimana penentuan distribusi posterior sulit dilakukan karena melibatkan

persamaan integral yang sangat kompleks. Untuk mendapatkan distribusi posterior parameter diperlukan proses integral dengan dimensi yang besar dan waktu yang cukup lama. Salah satu solusi untuk mengatasi masalah ini adalah dengan pendekatan numerik, yaitu *Markov Chain Monte Carlo* (Carlin dan Chib, 1995). Pada analisis Bayesian, penggunaan MCMC dapat mempermudah analisisnya, sehingga keputusan yang diambil dari hasil analisis akan dapat dilakukan dengan cepat dan tepat. Metode ini memiliki 2 algoritma yang sering digunakan yaitu algoritma *Metropolis-Hastings* dan algoritma *Gibbs Sampling*. (Walsh, 2004)

Algoritma Metropolis-Hastings merupakan algoritma untuk membangkitkan barisan sampel menggunakan mekanisme penerimaan dan penolakan (accept-reject) dari suatu distribusi posterior yang diinginkan untuk dilakukan penarikan sampel. Dalam algoritma MH dibutuhkan sebuah *proposal distribution* $p(\theta|\theta_{j-1})$ untuk membangkitkan kandidat sampel acak.

Algoritma Gibbs Sampling dapat dianggap sebagai kasus khusus dari Algoritma Metropolis-Hastings. *Gibbs Sampling* bisa diterapkan apabila distribusi probabilitas bersama (*joint probability distribution*) tidak diketahui secara eksplisit, tetapi distribusi bersyarat (*conditional distribution*) dari tiap-tiap variabel diketahui. Gibbs sampling dapat didefinisikan sebagai suatu teknik simulasi untuk membangkitkan variabel acak dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella dan George, 1992). Gibbs Sampling merupakan algoritma yang sering digunakan sebagai pembangkit variabel random dalam MCMC (Gelman dkk, 2014).

Untuk datanya, ada data tersensor dan ada pula data yang tidak tersensor. Untuk data tersensor terdiri dari data tersensor tipe I, tipe II, dan tipe III. Untuk analisis ini kita akan menggunakan data tersensor tipe I dengan lama waktu penelitian adalah satu tahun.



Gambar 2.6 Diagram Alir Kerangka Berpikir

BAB IV

PENUTUP

5.1 Kesimpulan

Berdasarkan analisis *survival* diperoleh estimasi parameter dari distribusi Lognormal dengan pendekatan Bayesian yaitu parameter $\mu = 1.881$ dan $\sigma = 3.415$. Parameter μ dan σ kemudian digunakan untuk mengetahui laju kesembuhan TB dengan menggunakan fungsi *hazard* dan fungsi *survival*. Perhitungan fungsi *survival* menunjukkan bahwa semakin lama pasien dirawat, maka kemampuan bertahan pasien akan semakin rendah dan semakin tinggi pula probabilitas pasien untuk menderita penyakit TB.

Dapat disimpulkan pula bahwa faktor-faktor yang berpengaruh secara signifikan terhadap waktu ketahanan hidup pasien rawat inap penderita TB di RSUD Tugurejo Semarang tahun 2017, adalah variabel umur. Dengan model regresi *cox proportional hazard* sebagai berikut:

$$\hat{h}(t) = h_0(t)\exp(0,009898x_1)$$

dimana $h_0(t)$ adalah *baseline hazard* yaitu mempunyai nilai yang sama dengan $h(t)$ jika variabel umur bernilai nol. Koefisien x_1 sebesar 0,009898, artinya semakin bertambah umur sebesar satu satuan maka penambahan nilai $\hat{h}(t)$ sebesar kelipatan $\exp(0,009898)$ satuan.

5.2 Saran

Berdasarkan hasil penelitian di atas, maka saran yang diberikan antara lain pada penelitian selanjutnya diharapkan menggunakan sampel yang lebih besar dan

rentang waktu yang lebih lama sehingga penelitian yang dihasilkan akan lebih baik. Kemudian untuk penelitian selanjutnya dengan kasus penyakit TB, pengambilan datanya sebaiknya tidak hanya dari rekam medis rumah sakit tetapi bisa dengan observasi langsung untuk lebih mendapatkan banyak informasi yang tidak tersedia di rekam medis misalnya status merokok dan keadaan rumah penderita. Analisis *survival* dapat dikembangkan dengan menggunakan pendekatan lain seperti, *non proportional hazard*, model *accelerated failure time* (AFT), model *cure*, model *frailty* atau model lain yang diharapkan dapat menambah cakupan penelitian selanjutnya.

DAFTAR PUSTAKA

- Amalia, S., Iriawan, N., & Dwi, P. (2013). Analisis Survival dan Faktor-Faktor yang Mempengaruhi Kesembuhan Pasien Demam Berdarah dengan Menggunakan Bayesian Mixture Survival. *Jurnal Sains dan Seni Its*.
- Bain, L. J. dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics Second Edition*. California: Duxbury Press.
- Bolstad, W. M. (2007). *Introduction to Bayesian Statistics 2nd Edition*. John Wiley: Canada.
- Box, G.E.P., & Tiao. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA : Addison-Wesley.
- Carlin, B. P dkk. (2010). Bayesian Methods for Data Analysis. *Journal of Applied Statistics Vol.37, No.4,705-706*.
- Collet, D. (2003). *Modeling Survival Data in Medical Research*. London : Chapman and Hall.
- Congdon, P. 2003. *Bayesian Statistical Modelling*. John Wiley: Chichester, UK.
- Cox, D. R. (1972). Regression Models and Life Tables (with discussion). *Journal of The Royal Statistical Society B, 34, 187-220*.
- Cox, D.R., Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Dahlan, S., (2012). *Analisis Survival : Dasar-Dasar Teori dan Aplikasi Program Stata*. Jakarta : Maumere.

- Depkes RI. (2002). *Pedoman penyakit tuberkulosis dan penanggulangannya*. Jakarta.
- Depkes RI. (2011). *Pedoman Nasional Penanggulangan Tuberkulosis*. Jakarta.
- Depkes RI. (2014). *Pedoman Nasional Pengendalian Tuberkulosis*. Jakarta, Depkes RI.
- Dwidayati, N. (2012). Analisis Cure rate Penderita Kanker Payudara Berdasar Pemodelan Regresi Cox. *SainteknoL*. Tersedia di <http://journal.unnes.ac.id/nju/index.php/sainteknoL/article/view/5554> [diakses 18-10-2016]
- Ernawatiningsih, N. P. L. (2012). Analisis Survival Dengan Model Regresi Cox. *Institut Teknologi Sepuluh November Vol. 2 No. 2*, 1693-1394.
- Gelman, A dkk. (1995). *Bayesian Data Analysis*. London : Chapman and Hall.
- Gilks, W. R., dkk. (1996). *Introducing Markov Chain Monte Carlo. Markov Chain Monte Carlo in Practice*. London : Chapman and Hall.
- Hanni, T., & Wuryandari, T. (2013). Model Regresi Cox Proporsional Hazard Pada Data Ketahanan Hidup. *Universitas Diponegoro Vol.6 No.1*, 11-20.
- Hiswani. 2006. TB merupakan Penyakit Infeksi yang Masih Menjadi Masalah Kesehatan Masyarakat. (On-line). www.dinkes-dki.go.id/tbl/html.
[Diakses 1 April 2011](#).

- Hosmer, D. W., Lemeshow, S. & May, S., 2008. *Applied Survival Analysis: Regression Modelling of Time to Event Data*. 2nd ed. New York: John Wiley and Sons. Inc..
- Ibrahim, J. G., dkk. (2000). Bayesian variable selection for proportional hazards models. *The Canadian Journal Of Statistics Vol. 28*, 1-20.
- Kementerian Kesehatan RI. 2015. *Profil Kesehatan Indonesia tahun 2014*. Jakarta: Kementerian Kesehatan RI.
- Kleinbaum, D. G., & Klein, M. (2005). *Survival Analysis A Self-Learning Text*. London : Springer.
- Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. Kanada : John Wiley & Sons, Inc.
- Lee, E. T., & Wang J. W. (2003). *Statistical Methods for Survival Data Analysis.3rd, ed.*New Jersey: John Wiley & Sons, Inc.
- Lee, E.T. 1992. *Statistical Methods for Survival Data Analysis*. John Willey and Sons, Inc. New York.
- Le, C. T., 1997. *Applied Survival Analysis*. New York: John Wiley and Sons. Inc.
- Lukitasari, D. (2015). Bayesian Survival Analysis untuk Mengestimasi Parameter Model Ketahanan Hidup Pasien Penderita Jantung Koroner. *JdC, Vol. 4, No. 1*.
- Mahmudah, N. (2014). Analisis Survival Weibull 3p Menggunakan Aplikasi Winbugs. *Universitas Brawijaya* 237-240.

- Misnadiarly. Mengenal, mencegah, menanggulangi TBC-Paru, Ekstra Paru pada anak dan kehamilan. OBOR, 2006.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Greece. John Wiley.
- Nyambega, H. O., dkk. (2017). Bayesian Estimation of Survivor Function for Censored Data Using Lognormal Mixture Distributions. *IOSR Journal of Mathematics (IOSR-JM) Vol. 13, No. 2*, 19-32.
- Putri, A. A., & Astutik, S. (2017). Survival Analysis On The Rate Of Diabetes Mellitus Patient Recovery With Bayesian Methode. *Proceeding of International Conference on Green Technology Vol.8 No.1*, 268-272.
- Rahajoe, dkk. 2005. Pedoman Nasional Tuberkulosis Anak. Jogjakarta: DIVA Press.
- Rejki, N. (2015). *Bayesian Reversible Jump Markov Chain Monte Carlo (Rjmc) untuk Pemodelan Mixture Survival*. Tesis. Surabaya : FMIPA Institut Teknologi Sepuluh Nopember.
- Sagita I., dkk. (2018). Perbandingan Metode Bayesian Self Dan Bayesian Linex Pada Estimasi Parameter Model *Survival* Distribusi Weibull Data Tersensor. *Buletin Ilmiah Math, Stat, dan Terapannya (Bimaster) Vol 07, No. 4*, 353-362.
- Sari, N. W. W., dkk. (2013). Model *Proportional Hazard Cox* Dengan Pendekatan Bayesian. *Jurnal Eksponensial Universitas Mulawarman Vol 4 No 1*, 9-16.

- Sari, O. (2016). *Analisis Life Time Penderita Demam Berdarah Dengue (DBD) Berdasarkan Asumsi Proportional Hazard (PH) Cox*. Skripsi. Semarang : FMIPA Universitas Negeri Semarang.
- Soejoeti, Z. Dan Soebanar. (1988). *Inferensi Bayesian*. Jakarta : Universitas Terbuka.
- Soemantri S, Bisara D, Philipus F, Indonesia Tuberculosis prevalence survey. 2004.
- Subanar. (2008). *Inferensi Bayesian*. Jakarta : Universitas Terbuka.
- Thamrin, S. A. (2008). Penggunaan Model Resiko *Proportional Cox* dengan Pendekatan Bayesian Semiparametrik menggunakan Prior Proses Gamma. *Paradigma, Vol 12 No.1*, 29-38.
- Thamrin, S. A., dkk. (2018). Penaksiran Parameter Distribusi Weibull Dengan Metode Bayesian Survival Dan Maksimum Likelihood. *Jurnal Keteknikan dan Sains (JUTEKS) – LPPM UNHAS Vol. 1 No. 2*, 22-27.
- Tolosie, K., & Sharma, M. K. (2014). Application of Cox Proportional Hazards Model in Case of Tuberculosis Patients in Selected Addis Ababa Health Centres, Ethiopia. *Tuberculosis Research and Treatment Vol 2014, Article ID 536976, 11 pages*.
- Walsh, B. (2004). *Markov Chain Monte Carlo and Gibbs Sampling*. Lecture Notes for EEB 581, version 26 April 2004.
- Wang, K., dkk (2017). Bayesian Cox Proportional Hazards Model in Survival Analysis of HACE1 Gene with Age at Onset of Alzheimer's Disease.

International Journal of Clinical Biostatistics and Biometrics Vol.3 No.1,
1-11.

Wijaya, A. A., & Wulandari, S. P. (2015). Analisis Survival pada Pasien Penderita Sindrom Koroner Akut di RSUD Dr. Soetomo Surabaya Tahun 2013 Menggunakan Regresi Cox Proportional Hazard. *Jurnal Sains dan Seni Its Vol. 4, No.2*, 2337-3520.

Wijayanti, R. (2014) *Perbandingan Analisis Regresi Cox Dan Analisis Survival Bayesian Pada Ketahanan Hidup Pasien Kanker Serviks Di Rsud Dr. Soetomo Surabaya*. Tesis. Surabaya : Institut Teknologi Sepuluh November.