Tesis – SS14 2501

PEMODELAN ANGKA LAMA SEKOLAH DI PROVINSI PAPUA BARAT DENGAN PENDEKATAN MODEL *MIXTURE SURVIVAL BAYESIAN*

MAULIDIAH NITIVIJAYA
NRP. 1314201709

DOSEN PEMBIMBING :
Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.
Dr. rer.pol. Heri Kuswanto, M.Si.

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

Thesis - SS14 2501

BAYESIAN SURVIVAL MIXTURE MODEL ON YEARS OF SCHOOLING IN WEST PAPUA PROVINCE

MAULIDIAH NITIVIJAYA
NRP. 1314201709

SUPERVISOR :
Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.
Dr. rer.pol. Heri Kuswanto, M.Si.

MAGISTER PROGRAM
DEPARTMENT OF STATISTCS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT yang telah melimpahkan karunia nikmat dan rahmat kepada penulis sehingga tesis yang berjudul "Pemodelan Angka Lama Sekolah di Provinsi Papua Barat dengan Pendekatan Model *Mixture Survival Bayesian*" ini dapat diselesaikan sesuai dengan waktu yang diharapkan. Tesis ini disusun dalam rangka memenuhi salah satu syarat untuk meraih gelar Magister Sains di Program Pasca Sarjana Statistika Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Badan Pusat Statistik (BPS) yang telah memberi kesempatan, dukungan dan beasiswa kepada penulis untuk melanjutkan studi program S2 di ITS.
2. Bapak Prof. Drs. H. Nur Iriawan, M.Ikom., Ph.D dan Dr. rer.pol. Heri Kuswanto, M.Si atas segala bimbingan dan arahannya dalam penyusunan tesis ini.
3. Bapak Dr. rer.pol. Dedy Dwi Prastyo, M.Si dan Ibu Dr. Margaretha Ari Anggorowati, M.T selaku dosen penguji yang telah memberikan saran dan koreksi atas penulisan tesis ini.
4. Bapak Dr. Suhartono, M.Sc selaku Kaprodi Pasca Sarjana Statistika ITS sekaligus dosen pembimbing akademik atas segala arahan dan bimbingan yang telah diberikan selama proses studi.
5. Bapak Ibu dosen Statistika ITS yang telah mencurahkan ilmu dan pengalamannya selama proses studi.
6. Seluruh jajaran staf jurusan Statistika ITS atas segala bimbingan dan fasilitas yang telah diberikan.
7. Suamiku tercinta, Anif Faisal, terima kasih atas segala pengertian, kesabaran, dukungan, nasehat, doa dan cintanya yang luar biasa diberikan kepada penulis sehingga penulis mampu menyelesaikan masa studi tepat waktu.
8. Permata hati dan penyejuk jiwaku, Hasna Kamila Alfaiz (4,5 tahun) atas kesabaran dan kerelaan menunggu umie dan abie pulang serta senantiasa

menjadi pelipur lara dengan segala canda tawanya. Doa dan harapan terbaik untuknya.

9. Ayah dan mama tersayang, atas segala doa, nasehat, waktu serta segala pengorbanan yang tak terhitung kepada penulis. Surga Allah tengah menanti kalian.
10. Ibu mertuaku atas doa dan nasehat yang diberikan, semoga penulis dapat mengambil hikmah yang berharga agar menjadi insan yang lebih baik.
11. Teman seperjuangan dan seatap, Mbak Afni atas segala keikhlasannya membantu penulis dalam segala hal, dan juga teman-teman angkatan 8 BPS yang tidak bisa disebutkan satu persatu atas bantuan dan kebersamaannya selama ini baik suka dan duka. Alhamdulillah bersyukur dapat mengenal kalian.
12. Teman-teman S2 Jurusan Statistika Kelas Reguler Angkatan 2014 serta teman-teman S3, Bu Eny, Mas Aji, Bu Suci, Bu Ani atas bantuan dan support-nya kepada penulis.
13. Semua pihak yang telah membantu penyelesaian tesis ini.

Surabaya, Januari 2015

Penulis

PEMODELAN ANGKA LAMA SEKOLAH DI PROVINSI PAPUA BARAT DENGAN PENDEKATAN MODEL *MIXTURE SURVIVAL BAYESIAN*

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)
di
Institut Teknologi Sepuluh Nopember
Oleh :

MAULIDIAH NITIVIJAYA
NRP. 1314201709

Tanggal Ujian : 26 Januari 2016
Periode Wisuda : Maret 2016

Disetujui Oleh :

1. Prof. Drs. Nur Iriawan, M.Ikom, Ph.D (Pembimbing I)
   NIP. 19621015 198803 1 002

2. Dr.rer.pol. Heri Kuswanto, M.Si (Pembimbing II)
   NIP. 19820326 200312 1 004

3. Dr.rer.pol. Dedy Dwi Prastyo, M.Si (Penguji)
   NIP. 19831204 200812 1 002

4. Dr. Margaretha Ari Anggorowati, M.T (Penguji)
   NIP. 19720222 199803 2 002

Direktur Program Pascasarjana

Prof. Ir. Djauhar Manfaat, M.Sc, Ph.D
NIP. 19601202 198701 1 001

KEMENTERIAN RISET, TEKNOLOGI DAN PENDIDIKAN TINGGI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
PROGRAM PASCASARJANA

PEMODELAN ANGKA LAMA SEKOLAH DI PROVINSI PAPUA BARAT DENGAN PENDEKATAN MODEL *MIXTURE SURVIVAL BAYESIAN*

Nama Mahasiswa	: Maulidiah Nitivijaya
NRP	: 1314201709
Pembimbing	: Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.
	Dr. rer.pol. Heri Kuswanto, M.Si.

ABSTRAK

Pendidikan merupakan salah satu pilar utama dalam menyusun suatu ukuran keberhasilan suatu wilayah. Indikator pendidikan angka lama sekolah menjadi salah satu target pemerintah dalam program Wajib Belajar 9 Tahun. Indikator ini menggambarkan betapa pentingnya pengetahuan dan keterampilan tingkat yang lebih tinggi. Provinsi Papua Barat sebagai salah satu provinsi termuda di Indonesia, diharapkan mampu bersaing mengembangkan kualitas sumber daya manusianya terutama di daerah tertinggal. Analisis data angka lama sekolah yang merupakan jenis data lama waktu, dalam statistika dikenal dengan analisis *survival*. Namun adakalanya dalam suatu penelitian yang melibatkan variabel respon ditemui adanya pola distribusi yang tidak mudah diamati sehingga menghasilkan model yang khas. Untuk itulah peneliti mencoba menerapkan model *mixture* pada angka lama sekolah. Estimasi model *mixture* dengan munculnya banyak parameter menimbulkan model yang kompleks sehingga digunakan pendekatan metode Bayesian melalui proses simulasi *Markov Chain Monte Carlo* (MCMC). Pembentukan model *mixture survival* dilakukan berdasarkan klasifikasi daerah tempat tinggal. Hasil penelitian menunjukkan bahwa data angka lama sekolah tersusun atas distribusi weibull pada masing-masing komponen dalam hal ini perkotaan dan perdesaan. Kontribusi yang dihasilkan dari distribusi mixture weibull angka lama sekolah yakni sebesar 40,12 persen di daerah perkotaan dan 59,88 di daerah perdesaan. Sedangkan peluang penduduk usia 16-24 tahun di daerah perkotaan dapat melanjutkan sekolah lebih tinggi dibandingkan di daerah perdesaan. Selain itu variabel yang berpengaruh terhadap angka lama sekolah di perkotaan juga berbeda dengan di perdesaan.

Kata kunci: *mixture Survival*, MCMC, regresi Cox, lama sekolah

BAYESIAN SURVIVAL MIXTURE MODEL ON YEARS OF SCHOOLING IN WEST PAPUA PROVINCE

By	:	Maulidiah Nitivijaya
Student Identify Number	:	1314201709
Supervisor	:	Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.
Co-Supervisor	:	Dr. rer.pol. Heri Kuswanto, M.Si.

ABSTRACT

Education could be considered as one of the basic pillars to determine the performance indicator of a respective region. Year of schooling is one of the education indexes, which becomes the government's target in the 9-year compulsory education program. This index illustrates the importance of knowledge and higher-level skills. Meanwhile, West Papua Province as one of the youngest provinces in Indonesia is challenged to improve the quality of human resources, particularly in the underdeveloped regions. Therefore, it is important to identify the variables which influence the years of schooling in the West Papua province. Statistically, the type of data such as length of time is frequently used to be the survival analysis. Nevertheless, the distribution pattern of the response variables is difficult to be analyzed. For that reason, this study applied mixture model on years of schooling. Mixture model estimation leads to the complex statistical problems with a number of parameters. Bayesian methods accomplish the estimation through the simulation process of Markov Chain Monte Carlo (MCMC). The survival mixture model was formed based on the status of county. Based on the results, the years of schooling are formed by weibull distribution for each component, which are rural and urban area. Rural areas were evidenced to give the contribution of years of schooling distribution more than urban area up to 59.88 percent. The opportunity to obtain formal education at high level school in rural areas was greater than urban area had. In general, the factors which influenced the years of schooling in urban and rural areas turned out to be different.

Key words: mixture *Survival*, MCMC, Cox regression, years of schooling

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

halaman ini sengaja dikosongkan

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Dalam upaya membangun sebuah bangsa yang maju dan modern sejatinya adalah melalui pendidikan. Selain sebagai sarana transfer pengetahuan dan keterampilan, pendidikan semestinya menjadi proses pembelajaran sepanjang hidup untuk membentuk karakter yang baik, mengembangkan potensi dan talenta individu serta menanamkan jiwa mandiri. Ketika pendidikan yang dimiliki oleh seseorang semakin tinggi maka kebudayaan dan peradaban yang unggul dan maju akan terbentuk.

Seiring dengan yang tertuang pada RPJMN 2015-2019 dimana salah satu agenda utama dalam bidang pendidikan yaitu melalui peningkatan taraf pendidikan penduduk. Wujud program yang ingin dicapai yakni pelaksanaan Wajib Belajar (Wajar) 12 Tahun yang ditujukan untuk mendorong pertumbuhan ekonomi yang berkesinambungan dan dalam upaya pengentasan kemiskinan. Karena pelaksanaan program tersebut mencakup keseluruhan proses pendidikan hingga siswa mampu menyelesaikan jenjang pendidikan menengah, maka berbagai permasalahan dalam pelaksanaan Wajar Pendidikan Dasar 9 Tahun yang belum terselesaikan harus dapat diatasi.

Pendidikan merupakan salah satu pilar utama dalam menyusun suatu indeks atau ukuran pembangunan di suatu wilayah atau dikenal dengan Indeks Pembangunan Manusia (IPM). Pada tahun 2013, indeks pendidikan pada penghitungan IPM di Indonesia masih diwakili oleh dua indikator yaitu Angka Melek Huruf dan rata-rata lama bersekolah (*mean years of schooling* / MYS). Angka melek huruf dapat diasumsikan sebagai langkah awal menuju dunia pengetahuan, sedangkan rata-rata lama sekolah menunjukkan betapa pentingnya pengetahuan dan keterampilan tingkat yang lebih tinggi. Namun pada tahun 2010, UNDP sudah merubah metodologi penghitungan IPM dengan metode baru dan direvisi pada tahun 2011. Penghitungan dengan metode baru ini mulai diimplementasikan di Indonesia pada tahun 2014. Salah satu tujuan perubahan metode tersebut yaitu agar indikator yang digunakan lebih tepat dan dapat membedakan dengan lebih baik. Salah satu yang dirubah yaitu penghitungan indeks pendidikan, dimana pada metode baru angka melek huruf tidak lagi digunakan. Cakupan indikator pendidikan yang digunakan pada metode baru yaitu rata-rata lama sekolah penduduk usia 25 tahun ke atas dan angka harapan lama sekolah (*Expected Years of Schooling*/EYS) penduduk usia 7 tahun ke atas.

Berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas) tahun 2014, capaian angka IPM di Indonesia masih tergolong dalam kategori sedang. Namun jika dilihat lebih dalam, masih terdapat provinsi dengan capaian IPM yang berada pada kategori rendah. Capaian IPM yang rendah salah satunya ditentukan oleh indeks pendidikan yang rendah pula. Papua Barat merupakan salah satu provinsi yang memiliki indeks pendidikan rendah, yakni berada pada urutan lima terbawah di level nasional. Rendahnya capaian ini bergantung pada besaran indikator pendidikan yang salah satunya adalah rata-rata lama sekolah penduduk. Rata-rata lama sekolah penduduk di Provinsi Papua Barat pada tahun 2014 hanya sebesar 8,66 yang artinya rata-rata penduduk di Provinsi Papua Barat belum mampu mengenyam pendidikan setingkat SMP. Angka ini tergolong cukup rendah jika dikaitkan dengan target program wajib belajar 9 tahun yang dicanangkan oleh pemerintah. Papua Barat yang merupakan salah satu provinsi termuda sangat disayangkan apabila pencapaian target pendidikan dasar belum terpenuhi. Dengan diberlakukannya Provinsi Papua Barat sebagai daerah otonomi khusus, pemerintah daerah diharapkan mampu mengembangkan kualitas sumber daya manusianya khususnya di daerah tertinggal.

Memperhatikan adanya target program peningkatan tingkat pendidikan, maka perlu dikaji tentang variabel-variabel apa yang mempengaruhi lamanya sekolah penduduk. Dengan mengetahui variabel tersebut diharapkan dapat memberikan gambaran sehingga dapat diambil kebijakan dalam rangka meningkatkan kualitas pendidikan di Provinsi Papua Barat.

Penelitian tentang rata-rata lama sekolah pernah dilakukan oleh Santoso (2009) yang memodelkan lama sekolah penduduk usia sekolah di Provinsi Papua dengan pendekatan *Spline Multivariable* dan MARS. Hasil penelitian tersebut diketahui bahwa variabel yang mempengaruhi lama sekolah adalah variabel topografi wilayah, umur anak, status pekerjaan anak, pendidikan kepala rumah tangga, pengeluaran perkapita, banyaknya anggota rumah tangga, lapangan pekerjaan kepala rumah tangga, dan status desa/kota. Setyawan (2011) juga melakukan pemodelan determinan tingkat pendidikan yang dilihat dari rata-rata lama sekolah dan angka melek huruf di Pulau Papua melalui pendekatan Regresi Nonparametrik *Birespon Spline*.

Penelitian lainnya dilakukan oleh Brunello dan Checchi (2005) yang meneliti tentang pengaruh rasio murid-guru dan tingkat pendidikan orang tua dikaitkan dengan tingkat pencapaian pendidikan penduduk di pasar tenaga kerja Italia pada tahun 1960-1980. Selanjutnya Solikhah (2009) melakukan analisis rata-rata lama sekolah di Pulau Kalimantan dengan menggunakan Model Spasial *Conditional Autoregression* (CAR) dimana unit observasi yang digunakan adalah wilayah kabupaten/kota.

Dalam statistika terdapat metode analisis *survival* dimana metode tersebut mempelajari lamanya suatu peristiwa atau kejadian yang biasa disebut *failure event*, misalnya pengamatan lama sekolah seseorang. Waktu dari awal perlakuan sampai terjadinya respon pertama kali yang ingin diamati disebut sebagai waktu ketahanan hidup (*survival time*) atau biasa disimbolkan T. Karena responnya berupa waktu, maka mungkin saja peristiwa yang diharapkan terjadi belum ditemukan pada saat pengumpulan data berakhir. Collet (1994) menyatakan bahwa pada kondisi demikian dikatakan sebagai pengamatan tersensor. Apabila *survival time* diamati dengan melibatkan variabel-variabel prediktor maka salah satu metode analisis yang sering digunakan adalah regresi *cox proportional hazard*.

Adakalanya dalam melakukan suatu penelitian yang melibatkan variabel lama waktu sebagai variabel respon, seringkali ditemui adanya pola distribusi data yang tidak mudah diamati dimana biasa tampak dari histogram data. Dari pola tersebut menghasilkan model yang khas dimana model ini akan tampak dari data yang diamati dan data yang ada biasanya terdiri dari beberapa sub populasi atau grup (McLachlan dan Basford, 1988). Model seperti ini jika diterapkan dalam analisis *survival* disebut model *mixture survival*. Analisis *survival* banyak diterapkan dalam bidang kesehatan, tetapi banyak juga dikembangkan pada bidang sosial. Muthen dan Masyn (2005) menggunakan analisis *mixture survival* yang salah satu aplikasi penelitiannya tentang kelangsungan sekolah dasar.

Estimasi model *mixture* dengan munculnya banyak parameter akan menimbulkan kesulitan tersediri sehingga pendekatan numerik atau metode *Bayesian* lebih mudah dilakukan melalui proses simulasi *Markov Chain Monte Carlo* (MCMC). Salah satu kelebihan model *mixture survival* dengan pendekatan metode *Bayesian* selain memperhitungkan pola distribusi data, dalam simulasi MCMC dapat menghasilkan estimasi parameter dengan lebih akurat. Hal ini dapat terjadi karena setiap parameter yang diestimasi memiliki distribusi marginal posterior tertentu sehingga nilai estimasi parameter yang dihasilkan akan lebih mendekati dengan nilai parameter yang sebenarnya. Hariyanto (2009) melakukan pemodelan lama mencari kerja di Pulau Jawa pada tahun 2007 berdasarkan daerah tempat tinggal menggunakan regresi *mixture survival bayesian.* Selanjutnya, analisis *survival* juga berkembang seperti penelitian yang dilakukan Qudsi (2015) yaitu memodelkan angka lama sekolah anak umur 16-18 tahun di Jawa Timur pada tahun 2012 dengan pendekatan model *mixture survival* spasial dimana efek random (*frailties*) diduga ikut berpengaruh.

Dalam kenyataannya, Provinsi Papua Barat yang terbagi atas 11 kabupaten/kota memiliki proporsi wilayah perdesaan yang jauh lebih banyak dibandingkan di wilayah perkotaan. Secara umum angka lama sekolah di Papua Barat masih tergolong rendah, dimana di wilayah tertentu pada usia sekolah menengah ke atas (SMA) mereka masih menduduki bangku SMP. Dalam kaitannya dengan rata-rata angka lama sekolah, di Provinsi Papua Barat cenderung ditemui adanya fenomena bahwa di daerah perkotaan rata-rata angka lama sekolah cenderung lebih tinggi dibandingkan di daerah perdesaan, yang artinya di daerah perdesaan anak usia sekolah rentan putus sekolah. Oleh karenanya akan lebih menarik apabila dalam penelitian ini dikaji faktor-faktor yang mempengaruhi angka lama sekolah di Papua Barat berdasarkan klasifikasi tempat tinggal pada anak dengan kategori usia 16-24 tahun. Dengan mempertimbangkan bentuk model *survival* terkait dengan variabel respon yang berupa waktu, serta adanya karakteristik distribusi angka lama sekolah yang khusus, maka dalam penelitian ini diusulkan model *Mixture survival Bayesian* pada angka lama sekolah di Provinsi Papua Barat berdasarkan klasifikasi daerah pada tahun 2014.

1.2 Perumusan Masalah

Dari uraian di atas, dapat diambil pokok permasalahan yang ingin diteliti yaitu:

1. Bagaimana cara memperoleh estimasi parameter *mixture survival Bayesian* berdasarkan klasifikasi daerah.
2. Bagaimana memperoleh model lama sekolah dengan pendekatan model *mixture survival Bayesian* berdasarkan klasifikasi daerah.
3. Bagaimana mendapatkan variabel-variabel yang berpengaruh terhadap lama sekolah di Provinsi Papua Barat menggunakan model *mixture survival Bayesian* berdasarkan metode *cox proportional hazard* untuk memberikan rekomendasi pada pemerintah daerah setempat.

1.3 Tujuan Penelitian

Tujuan penelitian ini yaitu:

4. Memperoleh cara estimasi parameter *mixture survival Bayesian* berdasarkan klasifikasi daerah.
5. Memperoleh model lama sekolah dengan pendekatan model *mixture survival Bayesian* berdasarkan klasifikasi daerah.
6. Mendapatkan variabel-variabel yang berpengaruh terhadap lama sekolah di Provinsi Papua Barat menggunakan model *mixture survival Bayesian* berdasarkan metode *cox proportional hazard* untuk memberikan rekomendasi pada pemerintah daerah setempat.

1.4 Manfaat Penelitian

Manfaat penelitian ini yaitu hasil penelitian diharapkan dapat dijadikan masukan kepada pemerintah khususnya pemerintah daerah Provinsi Papua Barat dalam menentukan kebijakan pembangunan, khususnya di bidang pendidikan.

1.5 Batasan Masalah

Berdasarkan ruang lingkup permasalahan di atas, maka batasan permasalahan dalam penelitian ini antara lain:

1. Unit observasi dalam penelitian ini adalah penduduk usia 16-24 tahun di Provinsi Papua Barat dengan status anak.
2. *Failure event* dalam analisis survival ini yaitu kejadian mendapatkan ijazah SMP/MTs sederajat. Sensor yang digunakan dalam penelitian ini adalah sensor kanan (*right censor*), yaitu apabila responden sampai masa pencacahan selesai belum mengalami *failure event* maka waktunya dibatasi hanya sampai dengan berakhirnya masa pencacahan.

BAB 2
TINJAUAN PUSTAKA

2.1. Pengujian Distribusi Data

Tahapan awal yang dilakukan dalam penelitian ini yaitu mengetahui distribusi data dari variabel respon melalui uji *goodness of fit.* Dalam analisis *survival* yang menjadi respon adalah data waktu dari suatu objek hingga terjadinya suatu kejadian tertentu. Pengujian *goodness of fit* dapat dilakukan dengan beberapa cara, diantaranya dengan menggunakan metode uji Anderson Darling, Kolmogorov-Smirnov, dan Chi-Square. Pada metode uji Kolmogorov-Smirnov fungsi distribusi kumulatif (CDF) empiris $F_{\mathrm{n}}(x_i)$ dibandingkan dengan fungsi distribusi hipotesis (CDF estimasi) sehingga statistik uji yang digunakan seperti berikut:

$$D_n = \sup\left|F_n(x) - \hat{F}(x)\right|, \tag{2.1}$$

dengan uji hipotesis adalah:

H_0 : data X merupakan variabel random independen yang berdistribusi sesuai dengan distribusi $\hat{F}(x)$ atau $D_n = 0$

H_1 : data X merupakan variabel random independen yang tidak berdistribusi sesuai dengan distribusi $\hat{F}(x)$ atau $D_n \neq 0$

H_0 akan ditolak jika $D_n > d_n$ atau p-value < alpha, dimana d_n adalah nilai yang diambil dari tabel Kolmogorov-Smirnov.

Selain itu, uji *goodness of fit* dapat dilakukan dengan metode Anderson Darling dengan statistik uji sebagai berikut:

$$A_n^2 = \left(-\frac{1}{n} \left\{ \sum_{i=1}^{n} (2_i - 1) \left[\ln F\left(x_i\right) + \ln(1 - F\left(x_{n+1-i}\right)) \right] \right\} \right) - n, \tag{2.2}$$

dimana F merupakan fungsi distribusi kumulatif dari distribusi yang dihipotesiskan dan x_i merupakan data waktu *survival* yang telah diurutkan. Dengan hipotesis yang sama dengan diatas, pengambilan keputusan tolak H_0 apabila $A_n^2 > a_{n,1-\alpha}$ atau *p-value* < α, dengan nilai $a_{n,1-\alpha}$ merupakan nilai tabel Anderson Darling. Data dikatakan mengikuti distribusi tertentu apabila nilai statistik Anderson-Darling semakin kecil (Iriawan dan Astuti, 2006).

2.2. Analisis *Survival*

Analisis *survival* adalah salah satu metode statistika untuk menganalisis data dimana variabel responnya berupa waktu sampai suatu peristiwa atau *event* terjadi. *Event* didefinisikan sebagai peristiwa ekstrim yang mungkin terjadi pada individu, misalnya kejadian sembuhnya seseorang maupun kematian seseorang setelah dilakukan proses pengobatan. Respon *survival* didefinisikan sebagai *range* waktu dari awal penelitian sampai suatu *event* terjadi atau sampai penelitian berakhir, misalnya *range* waktu awal penelitian sampai individu mati atau sampai penelitian berakhir (Kleinbaum dan Klein, 2005). Menurut Lee (1992), analisis *survival* lebih difokuskan untuk memprediksi peluang respon, *survival*, rata-rata waktu hidup (*life time*), mengidentifikasi resiko, serta memprediksi faktor-faktor yang berhubungan dengan respon.

Analisis *survival* banyak diterapkan dalam bidang medis dan biostatistika serta dikenal dengan berbagai istilah di bidang lain seperti: *event history analysis* dalam bidang sosiologi, analisis *reliability* atau *failure time analysis* dalam bidang teknik dan industri, dan *duration analysis* atau *transition analysis* di bidang ekonomi (Cox dan Oakes, 1984). Terdapat tiga elemen yang harus diperhatikan dalam menentukan waktu *survival* t (Zang, 2008), yaitu:

1. Waktu awal (*time origin/starting point*) yaitu titik awal objek mulai diteliti, misalnya tanggal dimulainya suatu pengobatan
2. *Failure time* yakni waktu berakhirnya *failure event* harus jelas, misalnya tanggal kematian atau tanggal keluar dari rumah sakit (telah dinyatakan sembuh)
3. Skala pengukuran waktu atau *measurement scale of time*, misal skala waktu yang digunakan dalam penentuan lama sekolah seorang anak, dalam tahun, bulan, atau lainnya

Analisis *survival* berbeda dengan analisis statistik lainnya, dimana pada pada analisis *survival* terdapat data tersensor. Tipe data lama waktu tidak dapat dianalisis menggunakan regeresi linier biasa karena terdapat data tersensor dan tidak pula dapat dianalisis dengan regresi logistik karena lamanya waktu data yang diikuti tidak sama, sehingga yang paling sesuai adalah menggunakan analisis *survival* (Vittinghoff, Glidden, Shiboski, dan McCulloch, 2005). Akibatnya dengan adanya data tersensor tersebut, Lee (1992) menyatakan bahwa analisis *survival* memiliki ciri khusus, yaitu distribusi data lama waktu (*life time*) bersifat menceng atau *skew*.

2.2.1. Data Tersensor

Data tersensor memungkinkan beberapa individu tidak bisa diobservasi secara penuh sampai terjadinya *failure event* (Miller, 1998). Collet (1994) berpendapat, secara umum terdapat tiga alasan terjadinya penyensoran, diantaranya sebagai berikut:

1. *Lost to follow up,* yaitu jika obyek pengamatan meninggal, pindah, atau menolak untuk berpartisipasi
2. *Drop out,* yaitu jika perlakuan harus dihentikan karena suatu alasan tertentu misalnya pemberian kemoterapi yang dihentikan karena efek buruknya lebih besar dibanding manfaatnya
3. *Termination of study* yaitu jika masa penelitian berakhir sementara obyek pengamatan belum mencapai pada *failure event*.

Collet (2003) menyatakan bahwa data tersensor dibagi menjadi tiga, yaitu:

1. Data tersensor kiri, terjadi bila waktu *survival* sebenarnya seorang individu kurang atau sama dengan waktu *survival* individu saat di observasi.
2. Data tersensor kanan, bila waktu *survival* seorang individu tidak lengkap sampai dengan saat periode *follow-up*. Dengan kata lain objek belum mengalami *failure event* hingga waktu pengamatan berakhir (*study end*).
3. Data Sensor interval, bila terjadi pemutusan dalam pengumpulan data dan obyek penelitian mengalami *failure* antar interval waktu tersebut.

Model *survival* digunakan untuk menjelaskan bagaimana resiko (*hazard*) terjadinya suatu *event* tertentu pada suatu waktu dipengaruhi oleh beberapa *covariate* berdasarkan teori yang menunjang peristiwa tersebut. *Hazard rate* adalah resiko sesaat suatu unit pengamatan pada suatu waktu tertentu yang bertahan, yakni tidak mengalami peristiwa yang dimaksud hingga waktu berakhir. *Baseline Hazard* merupakan resiko terjadinya suatu *event* atau kejadian tanpa mempertimbangkan adanya efek *covariate*, misalnya *time dependency* suatu peristiwa (Darmofal, 2008).

Dalam model semiparametrik *Cox*, tidak terdapat distribusi parametrik khusus untuk *baseline hazard*-nya. Akibatnya, model regresi *Cox* lebih mengacu hanya pada penggabungan informasi waktu peristiwa yang diamati dibanding dengan menentukan suatu distribusi tertentu untuk interval terjadinya suatu peristiwa. Model regresi *Cox* mengacu pada semiparametrik karena meskipun tidak ada bentuk distribusi tertentu yang digunakan untuk *baseline hazard* tetapi resiko terjadinya suatu peristiwa tetap dinyatakan sebagai fungsi dari *covariate*. Aksioma (2011) menyatakan bahwa kelebihan lain model semiparametrik *Cox yaitu* pada fleksibilitas model (berbagai bentuk *baseline hazard*).

2.2.2. Fungsi *Survival* dan Fungsi *Hazard*

Misalkan T adalah variabel random non negatif yang menggambarkan waktu *survival* individu dari suatu populasi. Peluang T pada analisis *survival* secara umum digambarkan ke dalam tiga fungsi yaitu fungsi kepadatan peluang (*probability density function*), fungsi *survival*, dan fungsi *hazard.* Jika T melambangkan waktu *survival* dan mempunyai distribusi peluang $f(t)$ maka fungsi distribusi kumulatif dinyatakan sebagai berikut:

$$F(t) = \Pr(T < t) = \int_0^t f(u)du \ , 0 < t < \infty. \quad (2.3)$$

Fungsi *survival* $S(t)$ dapat dinyatakan sebagai peluang seseorang dapat bertahan lebih lama dari suatu waktu t dan dinyatakan melalui persamaan berikut.

$$S(t) = \Pr(T \geq t) = 1 - \Pr(T < t) = 1 - F(t). \quad (2.4)$$

Fungsi *hazard* merupakan reaksi sesaat atau laju kegagalan (*failure*) sesaat ketika seseorang mengalami suatu *event* pada waktu ke-t dan dinyatakan sebagai berikut:

$$h(t) = \lim_{\Delta t \to 0} \left\{ \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right\} = \frac{f(t)}{S(t)}. \quad (2.5)$$

Untuk mengetahui hubungan antara fungsi *survival* dan fungsi *hazard,* maka digunakan teori probabilitas bersyarat. Pada teori probabilitas bersyarat, yaitu

$$P(A|B) = \frac{P(AB)}{P(B)},$$

sehingga pada persamaan (2.5) dapat ditentukan hubungannya yakni:

$$\frac{\Pr(t < T < t + \Delta t)}{\Pr(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{S(t)},$$

dengan *F(t)* adalah fungsi distribusi kumulatif dari *T*. Selanjutnya persamaan (2.5) dapat dituliskan menjadi:

$$h(t) = \lim_{\Delta t \to 0} \left\{ \frac{F(t+\Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)}.$$

Dengan mengambil turunan fungsi distribusi $F(t)$ didapatkan:

$$F'(t) = \lim_{\Delta t \to 0} \left\{ \frac{F(t+\Delta t) - F(t)}{\Delta t} \right\} = f(t),$$

maka diperoleh hubungan antara fungsi *survival* dan fungsi *hazard* yaitu sebagai berikut:

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

dengan *F(t)* = 1 – *S(t)* dan dapat dituliskan sebagai $\int_0^t f(t)\,dt = 1 - S(t)$. Apabila fungsi tersebut diturunkan terhadap *t* maka diperoleh

$$f(t) = \frac{d(1 - S(t))}{dt},$$

sehingga nilai *h(t)* menjadi:

$$h(t) = \frac{\left(\dfrac{d(1 - S(t))}{dt} \right)}{S(t)} = \frac{\left(-\dfrac{d}{dt} S(t) \right)}{S(t)} \text{ dan } -h(t)\,dt = \frac{d(S(t))}{S(t)}.$$

Kemudian fungsi di atas dintegralkan, maka didapatkan:

$$-\int_0^t h(t)dt = \int_0^t \frac{1}{S(t)} d(S(t)),$$

$$-\int_0^t h(t)dt = \ln S(t) \Big|_0^t = \ln S(t) - \ln S(0) = \ln S(t),$$

$S(t) = \exp\left[-\int_0^t h(t)\,dt \right]$, sehingga diketahui fungsi kumulatif *hazard* adalah

$$H(t) = \int_0^t h(t)\,dt. \qquad (2.6)$$

Hubungan antara fungsi kumulatif *hazard* yang dilambangkan *H(t)* dengan fungsi *survival* yang dilambangkan *S(t)* adalah

$$H(t) = -\ln S(t), \tag{2.7}$$

dimana t : Waktu yang diamati

T : Waktu *survival* seorang individu

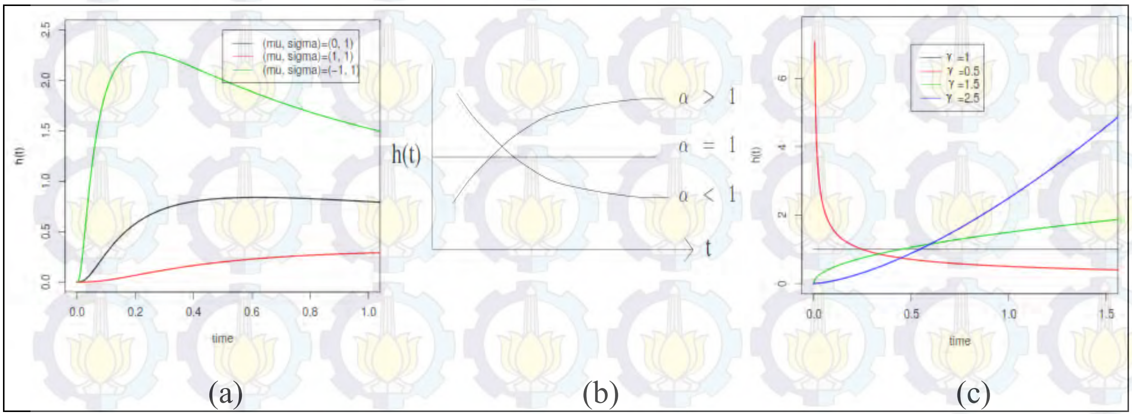
$F(t)$: Fungsi distribusi kumulatif dari distribusi data

$f(t)$: Fungsi kepadatan peluang (pdf)

$h(t)$: Fungsi *hazard* proporsional.

2.2.3. Pemodelan Fungsi *Hazard*

Secara umum, terdapat dua alasan dalam menentukan model data *survival*. Pertama, untuk menentukan kombinasi dari variabel prediktor yang paling berpotensi mempengaruhi fungsi *hazard* dan alasan kedua yaitu untuk mendapatkan estimasi fungsi *hazard* dari obyek itu sendiri. Fungsi *hazard* dengan notasi *h(t)* menyatakan laju kematian/kegagalan sesaat, yakni fungsi kegagalan jika suatu individu sudah dapat bertahan sampai waktu *t*. Fungsi ini dapat digunakan untuk membantu dalam pemilihan model sebaran data *survival time* (Lawless, 2003).



(a) (b) (c)

Gambar 2.1. Contoh Fungsi *Hazard*
(a) Distribusi Lognormal, (b) Distribusi Gamma, (c) Distribusi Weibull

Lee (1992) mendefinisikan fungsi *hazard* sebagai peluang kegagalan individu untuk bertahan selama interval waktu yang sangat pendek dengan asumsi bahwa individu tersebut telah bertahan pada awal interval atau limit peluang individu gagal bertahan dalam sebuah interval waktu yang sangat pendek, yaitu dari t sampai $t+\Delta t$ jika diketahui individu tersebut telah bertahan sampai waktu t seperti dituliskan pada persamaan (2.5). Semakin besar nilai *hazard* mengindikasikan bahwa resiko kegagalan yang dialami individu dalam penelitian semakin tinggi sehingga kemampuan bertahannya semakin kecil. Fungsi *hazard* dapat berupa fungsi naik, turun, konstan, atau menunjukkan fungsi yang lebih kompleks dan hal ini ditunjukkan pada Gambar 2.1.

Nilai variabel prediktor pada model *hazard* proporsional dinyatakan oleh vektor $\mathbf{x}$, dimana $\mathbf{x} = (x_1, x_2, ..., x_p)$. Fungsi *baseline hazard* dinyatakan sebagai $h_0(t)$ merupakan fungsi *hazard* untuk tiap-tiap individu dimana semua variabel prediktor dalam vektor $\mathbf{x}$ bernilai 0 (Collet, 2003). Fungsi *hazard* untuk obyek ke-i dapat ditulis sebagai:

$$h_i(t) = \psi(x_i) h_0(t),$$

dengan $\psi(x_i)$ adalah fungsi dari vektor variabel prediktor untuk obyek ke-i. Fungsi $\psi(x_i)$ dapat diinterpretasikan sebagai fungsi resiko seseorang pada waktu ke-t dengan vektor variabel prediktor x_i relatif terhadap resiko dari suatu obyek yang mempunyai $x = 0$. Adapun model umum *proportional hazard* adalah sebagai berikut:

$$h_i(t) = h_0(t)\exp(\beta_1 x_{1i} + \cdots + \beta_p x_{pi}). \tag{2.8}$$

Persamaan (2.8) disebut juga sebagai regresi *Cox*. Model tersebut dapat dinyatakan dalam bentuk lain, yaitu

$$\left[\frac{h_i(t)}{h_0(t)}\right] = \exp(\beta_1 x_{1i} + \ldots + \beta_p x_{pi}), \tag{2.9}$$

atau dapat digambarkan sebagai model linier dengan *log-relatif Hazard*, yakni

$$\ln\left[\frac{h_i(t)}{h_0(t)}\right] = (\beta_1 x_{1i} + \ldots + \beta_p x_{pi}). \tag{2.10}$$

Odds ratio dalam fungsi *hazard* adalah ukuran yang digunakan untuk mengetahui tingkat kecenderungan atau resiko, dengan kata lain merupakan perbandingan antara *odd* individu dengan kondisi variabel prediktor x pada kategori sukses dengan kategori gagal (Hosmer dan Lemeshow, 1999). Estimasi dari *odds ratio* didapatkan dengan mengeksponensialkan koefisien regresi *Cox* masing-masing variabel prediktor yang signifikan berhubungan dengan *hazard rate*-nya.

Misal X adalah variabel prediktor dengan dua kategori yaitu 0 dan 1. Hubungan antara variabel X dan $h(t)$ dinyatakan dengan $h_0(t)e^{\beta x}$ maka

untuk $x = 1$, fungsi *hazard* adalah

$$h(t|x=1) = h_0(t)e^{\beta.1} = h_0(t)e^{\beta}.$$

untuk $x = 0$, fungsi *hazard* adalah

$$h(t|x=0) = h_0(t)e^{\beta.0} = h_0(t).$$

Odds ratio untuk individu $x = 1$ dibanding $x = 0$ adalah

$$\frac{h(t|x=1)}{h(t|x=0)} = \frac{h_0(t)e^{\beta.1}}{h_0(t)e^{\beta.0}} = \frac{h_0(t)e^{\beta}}{h_0(t)} = e^{\beta}, \tag{2.11}$$

sehingga diperoleh nilai *odds ratio* yang artinya bahwa tingkat kecepatan terjadinya *failure event* pada individu dengan kategori $x = 1$ adalah sebesar e^{β} kali tingkat kecepatan terjadinya resiko terjadinya peristiwa *failure event* pada individu dengan kategori $x = 0$. Pada variabel kontinyu, nilai dari e^{β} mempunyai interpretasi perbandingan *odds ratio* antara individu dengan nilai x lebih besar 1 satuan dibanding individu lain (Hosmer dan Lemeshow, 1999).

2.2.4. Distribusi Weibull 2-Parameter

Bentuk fungsi densitas dari distribusi Weibull 2-Parameter dimana λ merupakan parameter skala dan γ merupakan parameter bentuk dapat ditulis sebagai berikut:

$$f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma}), \tag{2.12}$$

dengan nilai $\lambda > 0$, $\gamma > 0$.

Fungsi distribusi kumulatif dapat dinyatakan sebagai berikut:

$$F(t) = 1 - \exp(-\lambda t^{\gamma}). \tag{2.13}$$

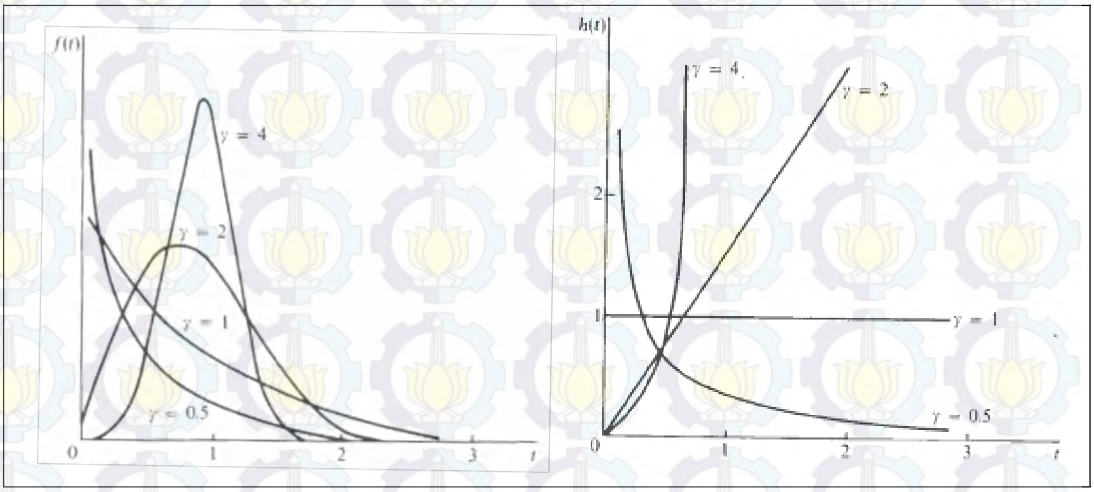
Berdasarkan persamaan (2.4) maka fungsi survival distribusi Weibull diperoleh

$$S(t) = \exp(-\lambda t^{\gamma}). \tag{2.14}$$

Selanjutnya bentuk fungsi hazard distribusi Weibull sesuai persamaan (2.5) dapat dinyatakan sebagai berikut:

$$h(t) = \lambda \gamma t^{\gamma-1}. \tag{2.15}$$

Gambar 2.2 menunjukkan kurva fungsi densitas dan fungsi hazard distribusi Weibull 2-parameter dengan nilai λ = 1 serta berbagai nilai γ yang berbeda.



Gambar 2.2. Kurva Fungsi Densitas (kiri) dan Fungsi *Hazard* (kanan) Distribusi Weibull dengan λ = 1

2.2.5. Asumsi *Hazard* Proporsional

Salah satu hal yang menarik dalam regresi *Cox* yaitu data tidak harus memenuhi distribusi apapun (Hosmer dan Lemenshow, 1999). Collet (1994) menyatakan bahwa asumsi pemodelan yang harus dipenuhi dalam regresi *Cox* yaitu asumsi *hazard* proporsional yang berarti fungsi *hazard* harus proporsional setiap waktu karena regresi *Cox* tidak mengakomodasi variabel yang berubah-ubah sepanjang waktu. Proporsional artinya variabel prediktor independen terhadap waktu dan hubungan antara *hazard* kumulatif sudah proporsional setiap waktu. Asumsi proporsional tersebut dapat diketahui dengan melihat pola plot –ln [–ln $S(t)$] atau ln [–ln $S(t)$] terhadap waktu *survival* untuk setiap variabel prediktor dengan skala kategorik seperti pada *odds ratio* di persamaan (2.11). Asumsi *hazard* proporsional terpenuhi jika pola plot antar kategori dalam variabel prediktor membentuk pola yang sejajar (Kleinbaum dan Klein, 2005). Pola yang saling berpotongan menunjukkan bahwa kategori antar variabel prediktor tidak memenuhi asumsi *hazard proportional*. Asumsi *proportional hazard* didasarkan pada fungsi probabilitas survival berikut:

$S(t,\mathbf{X}) = \left[S_0(t)\right]^{\exp\sum_{k=1}^{p}\beta_k X_k}$ yang bernilai $0 \leq S(t, \mathbf{X}) \leq 1$.

Jika diambil nilai logaritma fungsi tersebut maka menjadi:

$$\ln\left[S(t,x)\right] = \exp\left(\sum_{k=1}^{p}\beta_k X_k\right)\ln\left[S_0(t)\right].$$

Nilai logaritma dari $S(t, \mathbf{X})$ dan $S_0(t)$ akan bernilai negatif sehingga diberikan tanda negatif di depan logaritma yang selanjutnya dilakukan logaritma kembali. Secara matematis dapat ditulis sebagai berikut:

$$\begin{aligned}
\ln\left\{-\ln\left[S(t,\mathbf{X})\right]\right\} &= \ln\left[-\exp\left(\sum_{k=1}^{p}\beta_k X_k\right)\ln\left[S_0(t)\right]\right] \\
&= \ln\left[\exp\left(\sum_{k=1}^{p}\beta_k X_k\right)\right] + \ln\left(-\ln\left[S_0(t)\right]\right) \\
&= \sum_{k=1}^{p}\beta_k X_k + \ln\left(-\ln\left[S_0(t)\right]\right).
\end{aligned}$$

Atau dapat dituliskan,

$$-\ln\left\{-\ln\left[S(t,\mathbf{X})\right]\right\} = -\sum_{k=1}^{p}\beta_k X_k - \ln\left(-\ln\left[S_0(t)\right]\right). \tag{2.16}$$

Apabila mempertimbangkan dua spesifikasi dari vektor **X** pada dua individu yang berbeda yaitu $\mathbf{X}_1$ dan $\mathbf{X}_2$ dengan $\mathbf{X}_1 = (X_{11}, X_{12}, \ldots, X_{1p})$ dan $\mathbf{X}_2 = (X_{21}, X_{22}, \ldots, X_{2p})$, maka kurva log-log yang bersesuaian untuk individu tersebut yaitu dengan mensubstitusikan **X** dengan $\mathbf{X}_1$ dan $\mathbf{X}_2$ pada persamaan (2.16). Selanjutnya dihasilkan:

$$-\ln\left[-\ln S(t,\mathbf{X}_1)\right] = -\sum_{k=1}^{p}\beta_k X_{1k} - \ln\left[-\ln S_0(t)\right] \text{ dan}$$

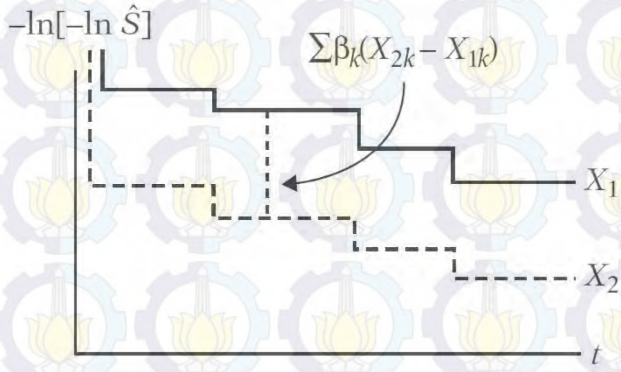
$$-\ln\left[-\ln S(t,\mathbf{X}_2)\right] = -\sum_{k=1}^{p}\beta_k X_{2k} - \ln\left[-\ln S_0(t)\right].$$

Dengan mengurangkan kurva log-log keduanya didapatkan hasil sebagai berikut:

$$-\ln\left[-\ln S(t,\mathbf{X}_1)\right] - \left(-\ln\left[-\ln S(t,\mathbf{X}_2)\right]\right) = \sum_{k=1}^{p}\beta_j\left(X_{2k} - X_{1k}\right), \tag{2.17}$$

dimana tidak mengandung unsur t atau independent terhadap t. Gambar 2.3 menunjukkan bahwa kurva plot $-\ln\ [-\ln\ S(t)]$ akan sejajar untuk setiap waktu t.

Sementara itu untuk variabel prediktor dengan skala *ratio* tidak memiliki asumsi apapun.



Gambar 2.3 Kurva Plot –ln [–ln $S(t)$] yang Sejajar

2.2.6. Estimasi Parameter Regresi *Cox Proportional Hazard*

Seperti halnya pada model regresi linier, pada regresi *Cox* kita perlu mengestimasi koefisien variabel prediktor $X_1, X_2, \ldots, X_p$ dalam komponen linier model, yaitu $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ sejumlah p variabel prediktor, maka estimasi fungsi *hazard* untuk objek ke-i dalam penelitian ini adalah:

$$\hat{h}_i(t) = e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i}\hat{h}_0(t), \tag{2.18}$$

dengan $\mathbf{x}_i$ adalah nilai vektor dari variabel prediktor untuk individu ke-i (i = 1, 2, ..., n) dan $\hat{h}_0(t)$ merupakan nilai estimasi dari fungsi *baseline hazard.* Estimasi fungsi *Hazard* dapat diperoleh jika nilai estimasi fungsi *baseline hazard* diperoleh terlebih dahulu.

Collet (1994) menyatakan bahwa koefisien $\boldsymbol{\beta}$ dalam model *proportional hazard* dengan nilai yang belum diketahui dapat diestimasi dengan menggunakan metode estimasi *partial likelihood*. Sebelum menjalankan metode ini, *likelihood* data sampel harus didapatkan dulu. Misal data untuk n sampel yang terdiri dari r waktu kejadian yang tidak tersensor dan n - r individu tersensor kanan, kemudian diurutkan menjadi $t_1 < t_2 < \cdots < t_j \ldots < t_n$ dengan $t_{(j)}$ merupakan urutan waktu *failure* ke-j. Diasumsikan terdapat r sampel *failure* dengan waktu yang berbeda dengan urutan waktu *failure* yang dinotasikan dengan $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$. Jika vektor variabel prediktor dari individu yang *failure* pada waktu $t_{(j)}$ dinotasikan dengan $\mathbf{x}_j$, maka peluangnya menjadi sebagai berikut:

P [individu dengan variabel $\mathbf{x}_j$ *failure* pada $t_{(j)}$ | satu *failure* pada $t_{(j)}$].

Misalkan kejadian A adalah individu dengan variabel $\mathbf{x}_j$ *failure* pada saat $t_{(j)}$ dan kejadian B adalah semua *failure event* pada saat $t_{(j)}$, maka

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)} \tag{2.19}$$

$$=\frac{P\left[\text{individu dengan variabel } x_{(j)}\, \textit{failure} \text{ pada } t_{(j)}\right]}{P\left[\text{semua } \textit{failure event} \text{ pada } t_{(j)}\right]}$$

Pembilang pada persamaan (2.19) adalah bentuk sederhana dari risiko *failure* pada waktu $t_{(j)}$ untuk individu dengan variabel $\mathbf{x}_j$. Apabila pembilang tersebut adalah individu ke-i yang mengalami kejadian *failure* pada saat $t_{(j)}$, maka fungsi *hazard* pada pembilang tersebut dapat ditulis menjadi $h_i(t_{(j)})$. Penyebut pecahan pada persamaan (2.19) adalah penjumlahan dari peluang *failure* pada waktu $t_{(j)}$ (yang dinotasikan $h_i(t_{(j)})$ dari semua individu yang mempunyai risiko *failure* pada waktu $t_{(j)}$ dimana individu-individu di dalamnya dinotasikan dengan indeks l.

Dalam suatu pengamatan, sampel-sampel yang beresiko *failure* pada saat $t_{(j)}$ dinyatakan dalam suatu himpunan R($t_{(j)}$), sehingga R($t_{(j)}$) dapat didefiniskan sebagai himpunan individu yang beresiko pada waktu $t_{(j)}$ yang terdiri dari individu-individu yang bertahan hidup hingga $t_{(j)}$. Peluang dalam persamaan (2.19) dapat ditulis menjadi:

$$P(A|B)=\frac{h_i\left(t_{(j)}\right)}{\sum\limits_{l\in R(t_{(j)})} h_l\left(t_{(j)}\right)}. \tag{2.20}$$

Selanjutnya diuraikan menjadi,

$$P(A|B)=\frac{h_0(t)\exp\left(\sum\limits_{k=1}^{p}\beta_k x_{ik}\right)}{\sum_{l\epsilon R\left(t_{(j)}\right)} h_0(t)\exp\left(\sum\limits_{k=1}^{p}\beta_k x_{lk}\right)},$$

$$P(A \mid B)=\frac{h_0(t)\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right)}{h_0(t)\sum_{l\epsilon R\left(t_{(j)}\right)}\exp\left(\sum_{k=1}^{p}\beta_k x_{lk}\right)},$$

sehingga didapat

$$P(A \mid B)=\frac{\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right)}{\sum_{l\epsilon R\left(t_{(j)}\right)}\exp\left(\sum_{k=1}^{p}\beta_k x_{lk}\right)}. \tag{2.21}$$

Cox (1972) menunjukkan bahwa fungsi *likelihood* yang sesuai untuk model *hazard* proporsional yaitu:

$$L(\boldsymbol{\beta} \mid \mathbf{x})=\prod_{j=1}^{r}\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{(j)})}{\sum_{l\in R(t_{(j)})}\exp(\boldsymbol{\beta}'\mathbf{x}_l)}, \tag{2.22}$$

dimana $\mathbf{x}_{(j)}$ adalah vektor variabel prediktor dari objek atau individu yang *failure* pada saat ke-j dengan urutan waktu $t_{(j)}$. Penjumlahan di penyebut pada persamaan (2.22) merupakan jumlah dari nilai exp($\boldsymbol{\beta}'\mathbf{x}_l$) dari seluruh objek yang beresiko pada waktu ke $t_{(j)}$.

Karena data yang diperoleh terdiri atas n pengamatan waktu *survival* dan ditunjukkan oleh $t_1, t_2, ..., t_n$, dengan δ_i adalah indikator event, yang bernilai nol untuk waktu *survival* ke-i dengan t_i (i = 1, 2, ..., n) merupakan tersensor kanan dan bernilai 1 untuk lainya, maka fungsi *partial likelihood* dapat dinyatakan dalam bentuk:

$$L(\boldsymbol{\beta} \mid \mathbf{x})=\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_l)}\right]^{\delta_i}, \tag{2.23}$$

dimana R(t_i) merupakan sekumpulan observasi pada waktu ke t_i dan fungsi log-*likelihood* yang bersesuaian dengan adanya sensor adalah sebagai berikut:

$$\ln L(\boldsymbol{\beta} \mid \mathbf{x})=\sum_{i=1}^{n}\delta_i\left(\boldsymbol{\beta}'\mathbf{x}_i-\ln\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_l)\right). \tag{2.24}$$

Dalam model *cox proportional hazard* estimasi parameter dapat diperoleh dengan cara memaksimumkan fungsi *likelihood.* Metode yang digunakan untuk mencari nilai estimasi parameter yakni melalui metode iterasi numerik Newton-Raphson, sehingga dengan diketahui fungsi log–*likelihood* diperoleh vektor $u(\boldsymbol{\beta})$ yang merupakan vektor berdimensi p x 1 dan merupakan turunan pertama fungsi log-*likelihood* terhadap parameter $\boldsymbol{\beta}$ pada persamaan (2.24)

$$u(\boldsymbol{\beta})=\frac{\partial \ln L(\boldsymbol{\beta}\,|\,\mathbf{x})}{\partial \boldsymbol{\beta}_w}=\sum_{i=1}^{n}\delta_i\left[\mathbf{x}_{wi}-\frac{\sum_{l\in R(t_i)}\mathbf{x}_{wl}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}{\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right].$$

dimana matrik $\mathbf{I}(\boldsymbol{\beta})$ adalah matrik informasi berukuran p x p dan merupakan matriks turunan negatif kedua dari fungsi log-*likelihood* sehingga elemen ke-(j,w) dari matrik $\mathbf{I}(\boldsymbol{\beta})$ adalah:

$$\mathbf{I}(\boldsymbol{\beta})=\frac{\partial^2 \ln L(\boldsymbol{\beta}\,|\,\mathbf{x})}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_w}$$

$$=-\sum_{i=1}^{n}\delta_i\left[\frac{\sum_{l\in R(t_i)}\mathbf{x}_{jl}\mathbf{x}_{wl}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}{\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right]-\left\{\left(\frac{\sum_{l\in R(t_i)}\mathbf{x}_{jl}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}{\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right)\left(\frac{\sum_{l\in R(t_i)}\mathbf{x}_{wl}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}{\sum_{l\in R(t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right)\right\}.$$

Estimasi parameter $\boldsymbol{\beta}$ pada iterasi ke-$(c+1)$, $\hat{\boldsymbol{\beta}}_{c+1}$ berdasarkan prosedur Newton-Raphson, adalah sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{c+1}=\hat{\boldsymbol{\beta}}_c+\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_c)u(\hat{\boldsymbol{\beta}}_c), \tag{2.25}$$

bagi setiap nilai c = 0, 1, 2, ..., dan agar menjamin konvergensi $\hat{\boldsymbol{\beta}}_{c+1}$ dengan $u(\hat{\boldsymbol{\beta}}_c)$ merupakan vektor dan $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_c)$ adalah invers matriks informasi. Proses diawali dengan menentukan nilai awal $\hat{\boldsymbol{\beta}}_0=\mathbf{0}$. Kemudian dilakukan proses iterasi $\hat{\boldsymbol{\beta}}_1=\hat{\boldsymbol{\beta}}_0-\mathbf{I}(\hat{\boldsymbol{\beta}}_0)^{-1}u(\hat{\boldsymbol{\beta}}_0)$. Proses akan berhenti apabila perubahan dalam fungsi log-*likelihood* kecil. Pada saat proses iterasi mencapai kondisi konvergen $\hat{\boldsymbol{\beta}}_{c+1}\cong\hat{\boldsymbol{\beta}}_c$, matrik varian-kovarian dari estimasi parameter dapat didekati dengan invers matriks informasi $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_c)$. *Standard error* dari nilai estimasi $\beta_1, \beta_2, \ldots, \beta_p$ diperoleh dari elemen diagonal matriks tersebut.

2.2.7. Pengujian Parameter Model

Setelah didapatkan variabel prediktor yang masuk dalam model, menurut Collet (1994) langkah selanjutnya adalah pengujian signifikansi parameter model dengan langkah-langkah sebagai berikut.

1. Uji Serentak

Pengujian ini dilakukan untuk melihat dari beberapa koefisien regresi yang didapat setidaknya ada satu saja yang signifikan. Adapun hipotesisnya sebagai berikut:

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

H_1 : minimal ada satu $\beta_k \neq 0$ dengan $k = 1, 2, \ldots, p$

p merupakan banyaknya variabel prediktor.

Statistik Uji yang digunakan yaitu:

$$G^2 = -2\ln\left[\frac{L_0}{L_p}\right], \tag{2.26}$$

dengan

$$L_p = \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{(j)})}{\sum_{l\in R(t_{ij})}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right]^{\delta_i} \text{ dan } L_0 = \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{(j)})}{\sum_{l(t_l\geq t_i)}\exp(\boldsymbol{\beta}'\mathbf{x}_{(l)})}\right]^{\delta_i},$$

keterangan:

L_0 = Nilai *likelihood* variabel bebas tereduksi

L_p = Nilai *likelihood* dengan semua variabel bebas

Statistik G^2 ini mengikuti sebaran *Chi-Square* dengan derajat bebas p.

Tolak H_0 jika $G^2 > \chi^2_{\alpha;p}$ atau *p-value* < α artinya bahwa variabel prediktor secara keseluruhan mempengaruhi variabel respon. H_0 ditolak berarti paling sedikit ada satu $\beta_k \neq 0$

2. Uji Parsial

Untuk menguji signifikansi masing-masing parameter (koefisien regresi β_k) secara parsial digunakan uji *Wald* (Collet, 1994). Adapun hipotesis sebagai berikut:

$H_0 : \beta_k = 0$

$H_1 : \beta_k \neq 0$ dengan $k = 1, 2, \ldots, p$

Statistik uji yang digunakan adalah uji Wald yang dirumuskan seperti di bawah ini:

$$W_k = \left[\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right]^2 \tag{2.27}$$

dengan keterangan:

$\hat{\beta}_k$ merupakan penduga β_k

$SE(\hat{\beta}_k)$ merupakan *standard error* dari β_k

Statistik *Wk* diasumsikan mengikuti sebaran *Chi-Square*. Hipotesis akan ditolak jika $W_k > \chi^2_{\alpha;1}$ atau *p-value* < α. H_0 ditolak berarti bahwa variabel prediktor ke-*k* secara parsial atau berdiri sendiri berpengaruh signifikan terhadap variabel respon.

2.3. Model *Mixture*

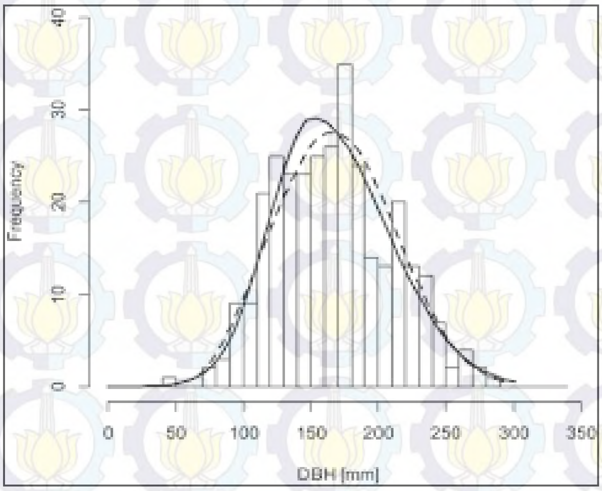
Model *mixture* merupakan suatu model yang khas dimana terlihat dari data yang diamati dan data yang ada terdiri dari beberapa subpopulasi atau grup. Setiap populasi yang terdiri dari beberapa subpopulasi maka suatu komponen dari *mixture* yang merupakan representasi distribusi dari sub populasi tersebut, dengan proporsi yang bervariasi untuk setiap komponennya (McLachlan dan Basford, 1988) dan (Gelman, dkk., 1995). Prinsip distribusi *mixture* yaitu menggabungkan sejumlah komponen yang mungkin berasal dari distribusi yang sama atau berbeda-beda sehingga dapat memberikan gambaran mengenai sifat-sifat data. Hasil distribusi *mixture* mampu memfasilitasi gambaran suatu sistem yang kompleks dengan lebih teliti. Marin dkk. (2001) menyatakan bahwa distribusi *Mixture* menyediakan kerangka parametrik yang fleksibel dalam permodelan dan analisis statistik.

Awal dikembangkannya model *mixture* yaitu berdasarkan penelitian Pearson dari data Weldon dalam McLachlan dan Peel (2000) yang mencoba melakukan *mixture* pada data berdistribusi normal yang berbeda μ_1 dan μ_2 serta variansi σ_1^2 dan σ_2^2 dalam proporsi π_1 dan π_2. Berbagai pendekatan yang telah dikembangkan pada model *mixture*, seperti algoritma *Expectation and Maximization* (EM), *Neural Network*, *Maximum Likelihood*, dan *Bayesian*.

Misalkan terdapat M komponen dalam sebuah *mixture*, diberikan $f_1(t)$, $f_2(t)$,..., $f_M(t)$ sebagai komponen densitas *survival* pertama, kedua sampai M komponen, maka model *mixture* yang disusun oleh M komponen tersebut akan dapat dituliskan :

$$f(t|\pi) = \pi_1 f_1(t) + \pi_2 f_2(t) + \cdots + \pi_M f_M(t), \tag{2.28}$$

dengan $\pi_1$adalah nilai kontribusi dari komponen *mixture* pertama, π_2 adalah nilai proporsi komponen *mixture* kedua, π_M merupakan nilai proporsi dari komponen *mixture* ke-M dan $\pi_1 + \pi_2 + \cdots + \pi_M = 1$. Model yang dijabarkan pada persamaan (2.25) disebut sebagai model *finite mixture* yang berlaku untuk model dengan jumlah komponen M tertentu.



Gambar 2.4. Pendekatan Model *Mixture* Normal dengan Dua Komponen (garis nyata) dan Model Plot Data (garis putus) (sumber: Pearson)

Menurut Stephen (1997), kemampuan model *mixture* sebagai metode analisis pada data yang mempunyai sifat campuran (*mixture*) menunjukkan keunggulan dibandingkan dengan metode statistik yang lain. Untuk mendeteksi kecenderungan apakah suatu data berdistribusi *mixture* dapat dilihat berdasarkan histogram data tersebut (Iriawan, 2001). Penelitian yang dilakukan Pearson, hasil model *mixture* akan lebih akurat hasilnya karena lebih mendekati distribusi data yang sebenarnya seperti yang dapat dilihat pada Gambar 2.4. Namun, Iriawan (2000) berpendapat, komponen *mixture* yang lebih banyak tidak selalu menjadikan kurva distribusinya lebih halus (*smooth*).

Estimasi distribusi *mixture* menggunakan metode *Bayesian* dilakukan dengan menemukan distribusi *posterior*, dengan cara mengalikan distribusi *prior* dan *likelihood* data. Selanjutnya, estimasi nilai setiap parameter modelnya dapat ditentukan setelah semua *prior* yang relevan telah diberikan (Gamerman, 1997). Estimasi model *mixture* dengan banyak parameter tentu akan memunculkan kerumitan tersendiri. Untuk menyelesaikan kesulitan ini, Iriawan (2000) berpendapat bahwa pendekatan *Bayesian* memiliki kelebihan dalam penarikan kesimpulan secara numerik, yaitu dengan menggunakan metode *Markov Chain Monte Carlo* (MCMC).

2.4. Analisis *Bayesian*

Dalam pendekatan *Bayesian*, data sampel yang diperoleh dari populasi, juga memperhitungkan suatu distribusi awal yang disebut *prior*. Menurut Iriawan (2001) untuk mengestimasi suatu parameter model data dikelompokkan menjadi dua bagian, yaitu data pengamatan saat ini yang bersifat sesaat selama studi dan data yang bersifat *long memory histogram*. Berbeda dengan pendekatan statistika klasik (*frequentist*) yang memandang parameter sebagai parameter bernilai tetap, pada pendekatan statistika *Bayesian* memandang parameter sebagai variabel random yang memiliki distribusi yang disebut sebagai distribusi *prior*. Dari distribusi *prior* selanjutnya dapat ditentukan distribusi *posterior* sehingga diperoleh estimator *Bayesian.*

Teorema *Bayesian* didasarkan pada distribusi *posterior* yang merupakan perpaduan antara distribusi *prior* (informasi masa lalu sebelum dilakukan observasi) dan data observasi yang digunakan untuk menyusun fungsi *likelihood* (Box dan Tiao, 1973). Hubungan distribusi *posterior* dengan distribusi *prior* dan *likelihood* dapat dituliskan sebagai berikut :

Distribusi *posterior* $\propto$ *likelihood* × Distribusi *prior*

Pada teorema Bayes, apabila terdapat parameter θ yang diberikan oleh data observasi *survival time* T, maka distribusi probabilitas untuk *posterior* θ pada data t akan proporsional dengan perkalian antara distribusi *prior* θ dan fungsi *likelihood* θ yang diberikan oleh data x. Secara matematis dapat dituliskan sebagai berikut:

$$f(\theta|t) = \frac{f(t|\theta)f(\theta)}{f(t)} \tag{2.29}$$

atau

$$f(\theta|t) \propto f(t|\theta)f(\theta). \tag{2.30}$$

Persamaan (2.21) $f(\theta|t)$ merupakan distribusi *posterior* yang proporsional dengan perkalian antara fungsi *likelihood* $f(t|\theta)$ dan distribusi *prior* $f(\theta)$. Menurut Iriawan (2001), jika diterapkan pada model *mixture* maka θ dapat disinkronisasi sebagai sebuah parameter yang memuat semua parameter model penyusun *mixture*. Spesifikasi dari disribusi *prior* sangat penting pada metode *Bayesian* karena distribusi *prior* mempengaruhi bentuk *posterior* yang akan digunakan untuk mengambil keputusan. Apabila informasi *prior* tersedia, maka informasi untuk distribusi *prior* akan terangkum didalamnya. Tetapi biasanya informasi *prior* tidak tersedia, sehingga perlu penetapan *prior* yang tidak akan mempengaruhi distribusi *posterior*. Distribusi tersebut biasa dikenal dengan sebutan *prior* sekawan (*conjugate*) yang parameterisasi distribusi *prior*nya tergolong sebagai *non-informative prior* atau *prior* samar-samar.

2.4.1. Distribusi *Prior*

Distribusi *prior* merupakan informasi yang terdahulu mengenai parameter. Pemilihan distribusi prior dalam pendekatan *Bayesian* harus tepat. Box dan Tiao (1973) menyatakan terdapat beberapa macam distribusi *prior* dalam metode *Bayesian*, antara lain:

1. *Conjugate prior* atau *non conjugate prior*, yaitu penentuan *prior* didasarkan pada pola *likelihood* dari datanya (Box dan Tiao, 1973)
2. Proper *prior* atau improper *prior* (Jeffreys *prior*) yaitu *prior* yang terkait dengan pemberian bobot atau densitas di setiap titik apakah terdistribusi secara *uniform* atau tidak. (Ntzoufras, 2009)
3. *Informative prior* atau *non informative prior* yaitu penentuan *prior* yang didasarkan pada ketersediaan pengetahuan atau informasi sebelumnya

mengenai pola distribusi data yang diperoleh dari penelitian sebelumnya. (Box dan Tiao, 1973)

4. *Pseudo prior* (Carlin dan Chib, 1995) menjabarkan penentuan *prior* dengan nilai yang disetarakan dengan hasil elaborasi cara *frequentist*, misalnya dengan priornya merupakan hasil dari estimasi parameter dengan metode maksimum *likelihood*.

2.4.2. Fungsi *Likelihood* pada Model *Mixture*

Menurut McLachlan dan Basford (1988), fungsi *likelihood* pada distribusi *Mixture* berbeda dengan fungsi *likelihood* pada distribusi univariat biasa. Apabila terdapat data pengamatan sebanyak n yang terdekomposisi ke dalam M kelompok data (subpopulasi) yang masing-masing mempunyai distribusi, maka fungsi *likelihood* model *mixture* tersusun dari beberapa *likelihood* data di setiap subpopulasi menurut distribusi masing-masing. Berdasarkan model *mixture* pada persamaan (2.21), fungsi *likelihood* model *mixture* adalah:

$$l_{mix} = \prod_{i=1}^{n} f_{mix}\left(t_i | \pi_m, \theta\right)$$

$$l_{mix} = \prod_{i_1=1}^{n_1} f\left(t_{i1} | \pi_1, \theta_1\right) + \prod_{i_2=1}^{n_2} f\left(t_{i2} | \pi_2, \theta_2\right) + \ldots + \prod_{i_m=1}^{n_M} f\left(t_{iM} | \pi_M, \theta_M\right), \qquad (2.31)$$

dengan syarat persamaan (2.28) adalah $n_1+n_2+ \ldots + n_M = n$ dan M adalah banyaknya komponen *mixture*.

2.4.3. *Markov Chain Monte Carlo* (MCMC)

Di dalam analisis *Bayesian*, penggunaan metode MCMC dapat mempermudah analisis sehingga keputusan yang diambil dari hasil analisis akan dapat dilakukan dengan cepat dan tepat. Menurut Carlin dan Chib (1995), pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks dan metode ini memungkinkan proses simulasi dengan mengambil sampel random dari model stokastik yang sangat rumit.

Ide dasar dari MCMC yakni membangkitkan data sampel dari distribusi *posterior* sesuai proses *markov chain* dengan menggunakan simulasi Monte Carlo secara iteratif sehingga diperoleh kondisi yang konvergen terhadap *posterior* (Ntzoufras, 2009). Kondisi seperti tersebut merupakan kondisi stasioner atau equilibrum. Selanjutnya, sampel parameter dalam *markov chain* diambil setelah kondisi stasioner tercapai sehingga sampel yang terambil dijamin merupakan sampel dari distribusi *posterior* dari parameter tersebut.

Iriawan (2000) berpendapat bahwa terdapat 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, estimasi densitas data dapat diketahui dengan cara membangkitkan suatu rantai markov yang berurutan sebanyak *n*.

2.4.4. *Gibbs sampling*

Terdapat beberapa teknik untuk memfasilitasi metode MCMC dalam mengestimasi parameter model, salah satunya adalah dengan *Gibbs sampler*. *Gibbs sampling* dapat didefinisikan sebagai suatu teknik simulasi untuk membangkitkan variabel random dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella dan George, 1992). *Gibbs sampler* merupakan generator yang sangat efisien sehingga sering digunakan sebagai generator variabel random pada analisis data yang menggunakan MCMC (Iriawan, 2000). Proses ini dilakukan dengan mengambil sampel dengan cara membangkitkan rangkaian *gibbs* variabel random berdasarkan sifat-sifat dasar proses *Markov Chain*. Dalam menjalankan program yang menggunakan rantai *markov* dilakukan pada kondisi bersyarat penuh. Ini merupakan salah satu kelebihan dari *Gibbs sampling* karena variabel random tersebut dibangkitkan dengan menggunakan konsep distribusi unidimensional yang terstruktur sebagai distribusi *full conditional*. *Gibbs sampling* sangat berguna dalam mengestimasi suatu parameter dalam suatu model yang kompleks yang mempunyai tingkat kerumitan dalam proses integritasi yang kompleks pula dan sulit diselesaikan secara analitis. Ilustrasi *Gibbs sampler* yang dikemukakan Casella dan George (1992) dapat dijelaskan pada contoh berikut.

Jika fungsi $f(t, \theta_{11}, \theta_{12}, \dots, \theta_{1p}, \theta_{21,}\theta_{22}, \dots, \theta_{2p}, \dots, \theta_{Mp})$ adalah *joint density* yang karakteristik densitas marginalnya untuk suatu θ_{11} dapat diperoleh dengan

$$f(\theta_{11}) = \int \dots \int f(t, \theta_{11}, \theta_{12}, \dots, \theta_{1p}, \theta_{21}, \dots, \theta_{2p}, \dots, \theta_{Mp}) d\theta_{11}, \theta_{12}, \dots, \theta_{1p}, \theta_{21}, \dots, \theta_{2p}, \dots, \theta_{Mp} \quad (2.32)$$

maka persamaan (2.32) diatas mungkin akan sulit untuk diselesaikan baik secara analitik maupun numerik. Metode *Gibbs sampler* memberikan alternatif untuk mendapatkan $f(\theta_{11})$ dengan cara membangkitkan sampel $\theta_{11}{}^{(1)}, \theta_{11}{}^{(2)}, \dots, \theta_{11}{}^{(w)}$ ~ $f(\theta_{11})$ tanpa membutuhkan $f(\theta_{11})$. Dengan melakukan simulasi sampel yang cukup besar, *mean*, *varians*, atau karakteristik apapun dari $f(\theta_{11})$ dapat dihitung dengan lebih tepat.

Ntzoufras (2009) menyatakan bahwa algoritma pada simulasi sampel dengan teknik *Gibbs Sampling* adalah sebagai langkah-langkah berikut:

1. Menentukan nilai awal untuk masing-masing parameter
   $\boldsymbol{\theta}^{(0)} = \theta_{11}{}^{(0)}, \theta_{12}{}^{(0)}, \dots, \theta_{1p}{}^{(0)}, \theta_{21}{}^{(0)}, \theta_{22}{}^{(0)}, \dots, \theta_{2p}{}^{(0)}, \dots, \theta_{Mp}{}^{(0)}$.
2. Untuk $c = 1, \dots, w$ ulangi langkah-langkah dibawah ini:
   a. menentukan $\boldsymbol{\theta} = \boldsymbol{\theta}^{(c-1)}$
   b. untuk $m = 1,\dots, M$ dan $k = 1,\dots, p$, update:
      θ_{mk} dari $\theta_{mk} \sim f(\theta_{mk} | \boldsymbol{\theta}_{\setminus mk}, \boldsymbol{t})$
   c. proses simulasi pada urutan pengambilan secara random setelah didapatkan nilai awal adalah sebagai berikut:

   $\theta_{11}{}^{(c)}$ dari $f(\theta_{11} | \theta_{12}{}^{(c-1)}, \dots, \theta_{1p}{}^{(c-1)}, \theta_{21}{}^{(c-1)}, \dots, \theta_{2p}{}^{(c-1)}, \dots, \theta_{Mp}{}^{(c-1)}, \boldsymbol{t})$

   $\theta_{12}{}^{(c)}$ dari $f(\theta_{12} | \theta_{11}{}^{(c)}, \theta_{13}{}^{(c-1)} \dots, \theta_{1p}{}^{(c-1)}, \theta_{21}{}^{(c-1)}, \dots, \theta_{2p}{}^{(c-1)}, \dots, \theta_{Mp}{}^{(c-1)}, \boldsymbol{t})$

   $\vdots$

   $\theta_{Mp}{}^{(c)}$ dari $f(\theta_{Mp} | \theta_{11}{}^{(c)}, \theta_{12}{}^{(c)} \dots, \theta_{1p}{}^{(c)}, \theta_{21}{}^{(c)}, \dots, \theta_{2p}{}^{(c)}, \dots, \theta_{M-1p-1}{}^{(c)}, \boldsymbol{t})$
3. Membentuk $\boldsymbol{\theta}^{(c)} = \boldsymbol{\theta}$ dan menyimpannya sebagai satu himpunan nilai yang dibangkitkan pada iterasi ke-(c + 1) dari algoritma.

2.5. Model *Mixture Survival*

Model *mixture* pada regresi merupakan pengembangan bentuk model *mixture* serta tidak bisa dilepaskan dari adanya *mixture* distribusi. Hurn, Justel, dan Robert (2000) melakukan pengembangan model *mixture* regresi yaitu *mixture* regresi logistik dan poisson pada data simulasi. Selain itu telah dilakukan pula *mixture* regresi linier pada kasus data IHK Indonesia (Iriawan, 2002).

Berdasarkan model persamaan (2.28), dapat dibentuk model *mixture* regresi *survival* dengan fungsi densitasnya tersusun dari distribusi *survival*nya. Dengan asumsi bahwa fungsi densitas tersebut adalah linier *mixture*, maka bentuk model *survival* dengan *M* komponen adalah sebagai berikut:

$$p(t|\pi,\theta)=\sum_{m=1}^{M} p(t|\theta_m,\pi_m). \tag{2.33}$$

Karena model *mixture* yang dibentuk sudah ditentukan di awal, yaitu berdasarkan klasifikasi daerah tempat tinggal responden yaitu perdesaan dan perkotaan, maka terdapat dua komponen *mixture*, sehingga persamaan yang akan dibentuk menjadi:

$$p(t|\pi,\theta)=\pi_1 p(t|\theta_1)+\pi_2 p(t|\theta_2), \tag{2.34}$$

dimana

$p(t|\theta_1)$: fungsi densitas untuk data *suvival* untuk daerah perkotaan
$p(t|\theta_2)$: fungsi densitas untuk data *suvival* untuk daerah perdesaan
π_1 : konstribusi untuk komponen distribusi *mixture* daerah perkotaan
$\pi_2 = (1-\pi_1)$: konstribusi untuk komponen distribusi *mixture* daerah perdesaan

2.6. Tinjauan Non Statistika

2.6.1. Partisipasi Sekolah

Keputusan untuk berpartisipasi dalam sekolah atau tidak oleh seseorang bukan hanya terjadi karena ketersediaan fasilitas pendidikan seperti ketersediaan kelas, tetapi juga karena adanya permintaan pendidikan dalam suatu rumah tangga atau individu dengan memperhitungkan pendapatan. Salah satunya, penelitian yang dilakukan oleh Schultz (2002) tentang keterkaitan antara tingkat pendidikan orang tua dengan kesehatan dan keinginan untuk menyekolahkan anak-anak mereka. McMahon (1999) menunjukkan hasil penelitian adanya hubungan yang positif antara pengeluaran rumah tangga pada bidang pendidikan terhadap angka partisipasi sekolah. Artinya, seorang anak bersekolah maupun tidak berkaitan erat dengan faktor ekonomi, sosial juga latar belakang keluarga.

Ketika pendapatan seseorang tinggi, maka memungkinkan seseorang menginvestasikan pendapatannya untuk pendidikan anak dan kesehatan keluarga. Namun ketika kekuatan perekonomian seseorang rendah, maka pemenuhan kebutuhan makanan cenderung lebih diutamakan dibandingkan investasi pendidikan. Keterbatasan ekonomi seringkali berimbas pada berbagai keterbatasan pemenuhan kebutuhan lainnya termasuk pendidikan. Terkait dengan hal itu, dengan mempertimbangkan banyaknya anggota rumah tangga maka seseorang akan memprioritaskan salah satu anggota rumah tangganya yang dianggap berpotensi untuk berpartisipasi dalam sekolah. Salah satu contohnya yaitu cara pandang dimana seseorang lebih mengutamakan anak laki-laki untuk bersekolah dibandingkan anak perempuan karena adanya pengaruh budaya. Ada pula cara pemikiran orang tua yang memutuskan bahwa anak-anak yang terpaksa harus berkorban untuk membantu mengangkat perekonomian keluarga dengan cara bekerja atau membantu pekerjaan orang tuanya.

Secara tidak langsung dari semua faktor yang ada, pola pemikiran orang tua sangat berpengaruh terhadap keputusan untuk menyekolahkan anak-anaknya. Pola pemikiran yang positif seringkali muncul karena tingkat intelegensi yang tinggi pula, artinya bahwa semakin tinggi tingkat pendidikan yang ditamatkan oleh seorang kepala rumah tangga maka semakin terbuka pula wawasan tentang pentingnya pendidikan bagi generasi penerusnya. Dengan tingkat pendidikan yang lebih tinggi maka diharapkan dapat mendapatkan pekerjaan yang lebih layak sehingga pemenuhan kesejahteraan keluarga menjadi semakin baik dan memutuskan rantai kemiskinan.

2.6.2. Penduduk Usia Sekolah

Penduduk usia sekolah menurut jenjangnya terbagi menjadi lima, antara lain sebagai berikut (BPS, 2006):

a. Usia 2-6 tahun, masuk dalam kelompok anak usia dini.

b. Usia 7-12 tahun, masuk dalam kelompok usia sekolah dasar (SD).
c. Usia 13-15 tahun, masuk dalam kelompok usia sekolah menengah pertama (SLTP).
d. Usia 16-18 tahun, masuk dalam kelompok usia sekolah menengah atas (SLTA).
e. Usia 19-24 tahun, masuk dalam kelompok usia sekolah di perguruan tinggi.

2.6.3. Angka Lama Sekolah dan Rata-Rata Lama Sekolah

Kemajuan suatu bangsa salah satunya diukur dari seberapa tinggi tingkat pendidikan rata-rata yang dimiliki oleh sebagian besar warganya. Semakin tinggi pendidikan seseorang maka semakin maju cara pandang seseorang dalam rangka mempertahankan diri untuk dapat meraih kehidupan yang lebih baik.

Lama sekolah didefinisikan sebagai lamanya pendidikan yang ditempuh seseorang dimana tidak memperhatikan cepat atau lambatnya seseorang dalam menempuhnya dari waktu yang telah ditargetkan. Jumlah tahun bersekolah ini tidak mengindahkan kasus-kasus tidak naik kelas, putus sekolah yang kemudian melanjutkan kembali, dan masuk sekolah dasar di usia yang terlalu muda atau sebaliknya, sehingga nilai dari jumlah tahun bersekolah menjadi terlalu tinggi kelebihan estimasi atau bahkan terlalu rendah (*underestimate*). Misalkan, seseorang yang menamatkan pendidikan tingkat SD artinya dia memiliki lama sekolah sebesar 6 tahun, jika mampu menamatkan hingga tingkat SMP artinya lama sekolah yang dia punya adalah 9 tahun.

Definisi ini sesuai dengan konsep BPS, dimana lama sekolah seseorang dihitung berdasarkan jenjang dan jenis pendidikan tertinggi yang pernah/sedang diduduki oleh seseorang. Artinya untuk menghitung angka lama sekolah dibutuhkan informasi berupa partisipasi sekolah, jenjang dan jenis pendidikan yang pernah/sedang diduduki, ijasah tertinggi yang dimiliki, dan tingkat/kelas tertinggi yang pernah/sedang diduduki. Rumus penghitungan angka lama sekolah berikut tahun konversinya sesuai BPS, UNDP (2004), dan Bappenas yaitu sebagai berikut:

$$ALS = \text{Tahun\_Konversi} + (\text{Kelas\_Tertinggi} - 1). \qquad (2.35)$$

Keterangan :

ALS : Angka Lama Sekolah

Kelas_Tertinggi : Tingkat / kelas yang terakhir diduduki
Tahun_Konversi : Tahun Pendidikan yang Ditamatkan
(SD = 6 tahun; SMP = 9 tahun; SMA = 12 tahun; D1 = 13 tahun; D2 = 14 tahun; D3 = 15 tahun; D4/S1 = 16 tahun; S2 = 18 tahun; S3 = 21 tahun)

Angka lama sekolah disini juga dapat digunakan sebagai evaluator pelaksanaan Program Wajib Belajar 9 Tahun yang dicanangkan oleh pemerintah, artinya untuk dapat mencapai target program tersebut maka angka lama sekolah harus sudah mencapai minimal 9 tahun.

Sementara itu, rata-rata lama sekolah (RLS) atau *Mean Years of Schooling* (MYS) didefinisikan sebagai jumlah tahun yang digunakan penduduk dalam menjalani pendidikan formal. Rata-rata lama sekolah di suatu daerah didefiniskan sebagai jumlah dari angka lama sekolah dari setiap warga di daerah tersebut dibagi dengan jumlah warga di daerah tersebut dengan rumus di bawah ini.

$$RLS = \frac{\sum_{i=1}^{n} ALS_i}{n}. \tag{2.36}$$

Keterangan :

RLS : Rata-rata Lama Sekolah suatu daerah
ALS : Angka Lama Sekolah
n : Jumlah penduduk suatu daerah

Rata–Rata Lama Sekolah menunjukkan rata-rata lamanya bersekolah seseorang dari masuk sekolah dasar atau sederajat sampai dengan tingkat pendidikan terakhir. Angka ini merupakan transformasi dari bentuk kategorik tingkat pendidikan tertinggi (TPT) menjadi bentuk numerik. Rata-rata lama sekolah seringkali dijadikan rujukan oleh pemerintah untuk melihat perkembangan pendidikan di daerah tersebut. RLS merupakan indikator yang telah ditetapkan oleh UNDP pada tahun 1990 untuk penyusunan IPM. Seiring dengan perkembangannya, obyek RLS yang menjadi acuan dalam penghitungan IPM ialah penduduk usia 25 tahun keatas.

Jika indikasi rata-rata lama sekolah ialah untuk melihat kualitas penduduk suatu daerah secara keseluruhan dalam mengenyam pendidikan formal, maka indikasi dari angka lama sekolah ialah untuk melihat kualitas individual setiap penduduk dalam mengenyam pendidikan formal.

Lamanya bersekolah merupakan ukuran akumulasi investasi pendidikan individu. Setiap peningkatan tahun sekolah diharapkan akan membantu meningkatkan pendapatan individu tersebut. Rata-rata lama bersekolah dapat dijadikan ukuran akumulasi modal manusia di suatu daerah. Tingginya angka dan rata-rata lama sekolah menunjukkan jenjang pendidikan yang pernah/sedang diduduki oleh seseorang. Dengan kata lain, semakin lama/tinggi jenjang pendidikan yang ditamatkan oleh individu semakin meningkat pula kualitas penduduk di suatu daerah. Karena rata-rata lama sekolah berasal dari perhitungan angka lama sekolah, maka untuk meningkatkan rata-rata lama sekolah suatu daerah pemerintah harus menaikkan angka lama sekolah setiap individu didaerah tersebut.

2.6.4. Variabel-Variabel Yang Mempengaruhi Angka Lama Sekolah

Berdasarkan yang telah diuraikan diatas, lama sekolah seseorang akan diperoleh dengan melihat partisipasi sekolah yang ditempuh. Banyak anak dalam kategori usia sekolah, tetapi karena sesuatu sebab tidak dapat bersekolah dan masih banyak pula anak bersekolah tidak dapat melanjutkan pendidikannya ke jenjang yang lebih tinggi. Penelitian Ersado (2005) tentang pekerja anak dan keputusan bersekolah di desa dan kota menyimpulkan bahwa ketersediaan sekolah yang bagus serta adanya upaya untuk mendapatkan pendidikan dan upah yang yang lebih tinggi akan berimbas pada intensitas pekerja anak dan kemungkinan seorang anak untuk tetap bersekolah. Dalam suatu artikel yang berjudul "Memotret yang Terlupakan dalam Era Pembangunan Otsus Papua" yang dimuat oleh Komunitas Pendidikan Papua (2009) menyebutkan bahwa banyak sekali ditemukan anak-anak usia sekolah di Papua yang seharusnya sekolah namun harus bekerja hanya karena memenuhi kebutuhan hidupnya.

Suendra (1999) menyatakan bahwa terdapat beberapa faktor yang menghambat seorang anak untuk menyelesaikan pendidikan dasar 9 tahun, di antaranya faktor ekonomi, faktor sosial budaya, faktor geografis, dan faktor lainnya. Faktor ekonomi seringkali dikaitkan dengan besarnya pendapatan yang dimiliki oleh suatu keluarga. Blanden dan Gregg (2004) dalam penelitiannya yang berjudul *Family Income and Educational Attainment*, menunjukkan bahwa terdapat korelasi antara latar belakang keluarga (yang diukur dengan tingkat pendapatan) dengan pencapaian pendidikan anak. Dalam kenyataannya, mengukur pendapatan suatu rumah tangga tidaklah mudah. Indeks yang mengukur tingkat kesejahteraan di Indonesia salah satu komponennya dihitung dengan pendekatan pengeluaran rumah tangga. Semakin besar pengeluarannya maka semakin banyak kebutuhan rumah tangga yang dapat terpenuhi, sehingga dapat dikatakan bahwa rumah tangga tersebut juga semakin sejahtera.

Selain faktor pengeluaran, Lloyd (1996) dalam penelitiannya di wilayah *Sub Saharan Africa* menunjukkan bahwa berbagai ukuran tingkat pendidikan memiliki hubungan yang negatif dengan banyaknya saudara kandung yang dimiliki (*family size*). Dengan kata lain, keluarga yang memiliki jumlah anggota lebih banyak maka pencapaian tingkat pendidikan yang berhasil ditempuh tidak lebih tinggi dibandingkan dengan keluarga yang memiliki jumlah anggota lebih sedikit.

Dengan memperhatikan penelitian-penelitian sebelumnya tentang angka lama sekolah, seperti penelitian Ikana (2005) dimana menyatakan bahwa jika dukungan ekonomi tidak memadai bagi kelangsungan pendidikan anak, maka anak perempuan sering kali mengalah pada anak laki-laki dan terpaksa putus sekolah, yang berarti jenis kelamin dari seorang anak bisa menjadi faktor tinggi rendahnya angka lama sekolah seorang anak.

Santoso (2009) dalam penelitiannya tentang determinan rata-rata lama sekolah penduduk di Provinsi Papua mendapatkan variabel prediktor yang mempengaruhi lama sekolah seseorang secara signifikan yaitu topografi wilayah, umur anak, status pekerjaan anak, pendidikan kepala keluarga, pengeluaran perkapita, banyaknya anggota rumah tangga, dan status desa/kota. Sementara itu Sulistiyawati (2009) dalam penelitiannya tentang faktor yang mempengaruhi angka lama sekolah di daerah kota dan desa di kabupaten Boalemo Provinsi Gorontalo, mengungkapkan bahwa faktor umur, status pekerjaan, pendidikan kepala rumah tangga, dan pengeluaran rumah tangga merupakan faktor-faktor yang mempengaruhi angka lama sekolah untuk responden yang bertempat tinggal di daerah perkotaan, sedangkan faktor jenis kelamin, umur, status pekerjaan, pendidikan kepala rumah tangga, pengeluaran rumah tangga, dan banyaknya anggota rumah tangga.

Pusat Penelitian dan Pengembangan Kependudukan BKKBN (2011) melakukan kajian tentang profil penduduk remaja dimana disebutkan bahwa usia remaja merupakan fase umur penduduk yang sangat menentukan kualitas penduduk pada masa depan. Apabila usia remaja memperoleh pendidikan formal dan non formal yang cukup maka kualitas penduduk yang bersangkutan pada fase umur dewasa akan cenderung lebih baik sehingga menghasilkan generasi yang berkualitas. Delprato, dkk (2015) membahas tentang dampak pernikahan muda pada *Output* pendidikan di beberapa negara Sub-Saharan Afrika dan Asia Barat dimana dalam penelitian tersebut menunjukkan adanya hubungan antara menunda pernikahan dini dengan peningkatan lamanya menempuh pendidikan. Adanya faktor ekonomi dan sosial budaya mendorong terjadinya pernikahan di usia dini. Qudsi (2015) melengkapi penelitian tentang lama sekolah di Provinsi Jawa Timur dengan memasukkan faktor status perkawinan juga pengaruh spasial.

Hiliry (1995) dalam penelitiannya di Brazil menyebutkan bahwa faktor residensi desa atau kota dan tingkat kesejahteraan keluarga merupakan faktor yang paling dominan mempengaruhi apakah seorang anak melanjutkan sekolah atau tidak, dan bekerja atau tidak. Klasifikasi wilayah perkotaan dan perdesaan seperti halnya konsep di BPS, mengacu pada karakteristik sosial ekonomi, kondisi dan sejumlah akses ke fasilitas perkotaan, serta ciri dan tipologi lingkungan.

Berdasarkan penjelasan di atas pada penelitian ini variabel-variabel prediktor yang diduga mempengaruhi angka lama sekolah anak umur 16-24 tahun berdasarkan klasifikasi daerah perkotaan dan perdesaan yaitu jenis kelamin, status perkawinan, status bekerja anak, tingkat pendidikan kepala rumah tangga, jumlah anggota rumah tangga, rata-rata pengeluaran rumah tangga per kapita sebulan.

halaman ini sengaja dikosongkan

BAB 3
METODOLOGI PENELITIAN

3.1. Sumber Data

Sumber data yang digunakan dalam penelitian ini yaitu data sekunder bersumber dari *raw data* Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2014 di Provinsi Papua Barat. Data yang diambil adalah data angka lama sekolah dan responden yang dianggap sebagai *failure event* yaitu apabila responden lulus pada jenjang pendidikan setingkat sekolah menengah pertama (SMP) baik dengan status sekolah negeri maupun swasta serta berhasil memiliki ijazah pada rentang umur 16-24 tahun.

3.2. Metode Pengumpulan Data

Susenas merupakan survei dengan unit sampel rumah tangga terbesar yang dilaksanakan oleh BPS dengan lingkup nasional. Susenas 2014 menggunakan tiga jenis kerangka sampel dengan urutan tahapan, yaitu: kerangka sampel untuk pemilihan kecamatan, pemilihan blok sensus (BS) dari kecamatan terpilih, dan pemilihan sampel rumah tangga (ruta) dari BS terpilih. Masing-masing rumah tangga terpilih dikunjungi oleh petugas pencacah dari BPS (dalam hal ini Koordinator Statistik Kecamatan yang biasa disebut KSK atau staf BPS Kabupaten/Kota setempat) yang bertugas mewawancarai responden sesuai dengan daftar pertanyaan dalam kuisioner yang sudah disiapkan. Selain itu petugas terlebih dahulu dibekali dengan konsep dan definisi dalam bentuk pelatihan atau refreshing serta simulasi survei. Wawancara dilakukan langsung terhadap kepala rumah tangga atau kepada anggota rumah tangga yang dianggap mengetahui keadaan di rumah tangga tersebut.

Adapun jumlah sampel terpilih dalam Susenas tahun 2014 di Provinsi Papua Barat adalah sebanyak 15.927 responden. Namun yang digunakan dalam penelitian ini adalah responden (penduduk) usia 16-24 tahun ke atas yang berstatus anak.

Selanjutnya pengklasifikasian perkotaan dan perdesaan yang dipergunakan dalam penelitian sesuai dengan konsep yang digunakan pada Peraturan Kepala

(Perka) BPS Nomor 37 Tahun 2010 tentang klasifikasi perkotaan dan perdesaan di Indonesia. Definisi perkotaan adalah suatu wilayah administratif setingkat desa/kelurahan yang memenuhi persyaratan tertentu dalam hal kepadatan penduduk, persentase rumah tangga yang bergerak di sektor pertanian dan skoring tersedianya fasilitas baik sarana pendidikan formal, sarana kesehatan umum, pasar, jalan aspal, listrik, dan sebagainya. Begitu sebaliknya perdesaan adalah suatu wilayah administratif setingkat desa/kelurahan yang belum memenuhi persyaratan tertentu dalam hal kepadatan penduduk, persentase rumah tangga yang bergerak di sektor pertanian dan skoring tersedianya fasilitas baik sarana pendidikan formal, sarana kesehatan umum, pasar, jalan aspal dan listrik, dan sebagainya.

3.3. Variabel Penelitan

Berikut merupakan variabel-variabel yang akan dipergunakan sebagai variabel respon dan prediktor dalam penelitian beserta definisinya.

3.3.1. Variabel Respon

Variabel respon yang digunakan dalam penelitian ini adalah data lama sekolah anak usia 16-24 tahun pada tahun 2014 yaitu seseorang pada usia tersebut (yang tidak pernah menempuh sekolah, yang sedang bersekolah, dan yang pernah menempuh sekolah namun kini sudah tidak bersekolah lagi) berada dalam periode penelitian (tahun 2014) dalam satuan tahun dengan ketentuan sebagai berikut:

a) Jika seorang responden pernah bersekolah ataupun masih bersekolah hingga dinyatakan lulus karena mendapatkan ijazah sekolah menengah pertama (SMP) atau yang sederajat (MTs dan paket B) baik negeri maupun swasta dan dalam batas periode penelitian, maka waktu *survival* dikategorikan sebagai data *survival* tidak tersensor
b) Jika seorang responden sampai dengan batas periode penelitian, terhitung dari responden tersebut mulai bersekolah (tidak dibatasi kapanpun responden tersebut bersekolah) hingga responden tersebut mengalami hal-hal berikut:

1. Masih bersekolah pada jenjang SD/sederajat atau SMP/sederajat atau melebihi batas akhir penelitian (tahun 2014),
2. Jika seorang responden tidak pernah sama sekali bersekolah,
3. Jika seorang responden keluar dari masa belajar (mengundurkan diri) atau mengalami *drop out* (dikeluarkan dari sekolah),

maka data *survival* tersebut dikatakan data tersensor.

Variabel respon dikategorikan menjadi:

t=0, data tersensor jika responden tidak mengalami *failure event*. Misal adanya responden tidak pernah bersekolah sama sekali, keluar dari masa belajar (mengundurkan diri) atau mengalami *drop out* (dikeluarkan dari sekolah), atau melebihi batas akhir penelitian dan responden masih belum mendapatkan ijazah SMP/sederajat.

t=1, data tidak tersensor jika responden mendapatkan ijazah SMP/sederajat yang berada dalam batas periode penelitian.

3.3.2 Variabel Prediktor

Variabel prediktor yakni variabel yang digunakan untuk memprediksi variabel respon. Adapun variabel prediktor yang digunakan pada penelitian ini sebanyak enam variabel, yaitu:

a. Jenis Kelamin Responden (X_1), adalah jenis kelamin dari responden. Variabel jenis kelamin dikategorikan "1" untuk laki-laki dan "2" untuk perempuan.
b. Status Bekerja (X_2), yaitu status bekerja responden saat pencacahan. Kategori dari status bekerja adalah bekerja dengan kode "1" dan tidak bekerja dengan kode "2".
c. Status Perkawinan Responden (X_3), yaitu status perkawinan responden saat pencacahan. Status perkawinan yang biasa dikenal adalah belum kawin, kawin, cerai hidup, atau cerai mati. Responden yang berstatus kawin, cerai hidup, atau cerai mati. Selanjutnya dalam penelitian ini diberikan pengkodean ulang menjadi kawin, sehingga kode status perkawinan adalah "1" untuk tidak kawin dan "2" untuk kawin.

d. Tingkat Pendidikan Kepala Rumah Tangga (X_4) yaitu tingkat pendidikan tertinggi yang dimiliki oleh seorang kepala rumah tangga. Variabel berskala ordinal, dan pengkategorian variabel ini sebagai berikut : “1” untuk tidak tamat SD, “2” untuk tamat SD, “3” untuk tamat SMP, “4” untuk tamat SMA ke atas.
e. Jumlah ART (X_5) adalah jumlah anggota rumah tangga yang berada dalam satu rumah tangga.
f. Rata-rata pengeluaran rumah tangga perkapita sebulan (X_6) yaitu rata-rata pengeluaran baik makanan maupun non makanan dalam sebulan yang dikeluarkan oleh rumah tangga untuk setiap orang dalam satuan juta rupiah.

3.4. Metode dan Tahapan Penelitian

Sebelum melakukan tahapan penelitian, terlebih dahulu dilakukan tahap *pre-processing* data yang akan diolah. Tahapan persiapan data dan pemilihan variabel adalah sebagai berikut:

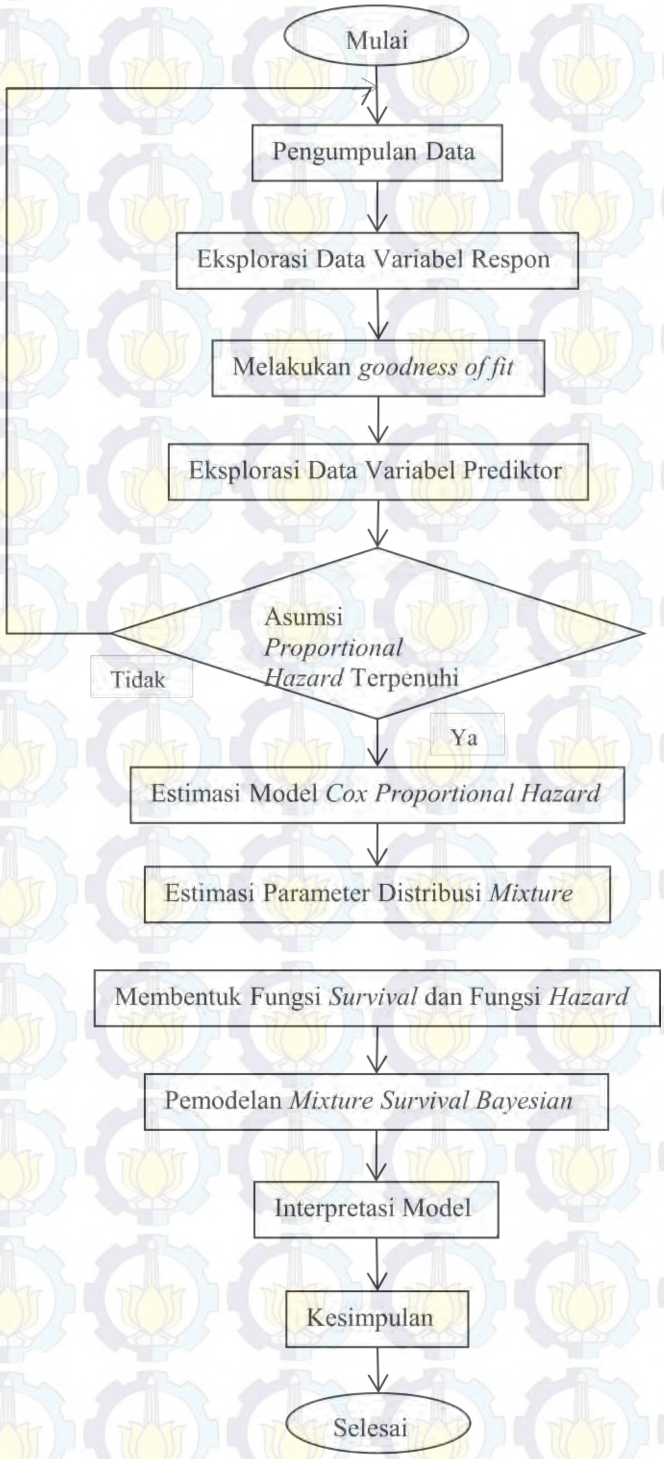
1. Menggabungkan variabel Susenas 2014 Kor individu dan Susenas 2014 Kor rumah tangga. Dari proses ini akan diperoleh informasi rata-rata pengeluaran perkapita per bulan (X_6).
2. Mengidentifikasi jumlah anggota rumah tangga dan tingkat pendidikan ART serta masing-masing kepala rumah tangga (KRT) pada setiap responden sehingga didapatkan variabel tingkat pendidikan KRT (X_4) dan jumlah ART (X_5) untuk setiap responden.
3. Menghitung angka lama sekolah penduduk berdasarkan persamaan (2.30) dengan range usia penduduk 16-24 tahun dan berstatus anak.
4. Menggabungkan variabel-variabel penelitian yang bersesuaian ke dalam satu set data.
5. Membentuk data yang diklasifikasikan menurut klasifikasi daerah yaitu perkotaan dan perdesaan.

Selanjutnya metode dan tahapan penelitian yang akan dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

a. Identifikasi variabel yang mempunyai efek signifikan atau yang berpengaruh terhadap lama bersekolah penduduk di Provinsi Papua Barat dan klasifikasi daerah tempat tinggal dengan menggunakan regresi *Cox proportional hazard*.
   1. Eksplorasi data variabel lama sekolah menurut daerah perkotaan dan perdesaan,
   2. Melakukan *goodness of fit* masing-masing distribusi data lama sekolah menurut pedesaan dan perkotaan,
   3. Eksplorasi data variabel prediktor (*X*) menurut pedesaan dan perkotaan di Provinsi Papua Barat,
   4. Estimasi parameter model regresi *Cox proportional hazard* untuk wilayah perkotaan dan perdesaan dan estimasi parameter distribusi *mixture* yang terbentuk. Adapun langkah-langkah yang dilakukan tahapan ini adalah sebagai berikut:
      i. Melakukan uji *proportional hazard* pada variabel prediktor,
      ii. Melakukan uji serentak untuk parameter model regresi *Cox proportional hazard,*
      iii. Melakukan uji parsial untuk parameter model regresi *Cox proportional hazard*.
b. Melakukan estimasi parameter distribusi *mixture* berdasarkan distribusi data (posterior) yang diperoleh pada pengujian *goodness of fit* dengan MCMC melalui iterasi *Gibbs Sampler*.
c. Memeriksa konvergensi algoritma MCMC dari estimasi parameter model *mixture* yang dihasilkan.
d. Berdasarkan hasil estimasi parameter regresi *Cox proportional hazard* pada langkah 4, nilai parameter tersebut digunakan sebagai dasar pembentukan distribusi *prior* dengan menggunakan *pseudo prior* serta hasil estimasi parameter model *mixture* pada langkah (b) sebagai dasar pembentukan distribusi prior untuk menyusun dan mengestimasi *mixture* regresi *survival* pada data lama sekolah. Penyusunan model *mixture* didasarkan pada klasifikasi daerah perkotaan dan perdesaan.
e. Memeriksa konvergensi algoritma MCMC dari estimasi parameter model *mixture proportional hazard* yang dihasilkan.

f. Interpretasi model *mixture proportional hazard* berdasarkan hasil estimasi parameter yang signifikan pada daerah perkotaan dan perdesaan.

3.5. Diagram Alir



3.6. Kerangka Pemikiran Penelitian

halaman ini sengaja dikosongkan

BAB 4
HASIL DAN PEMBAHASAN

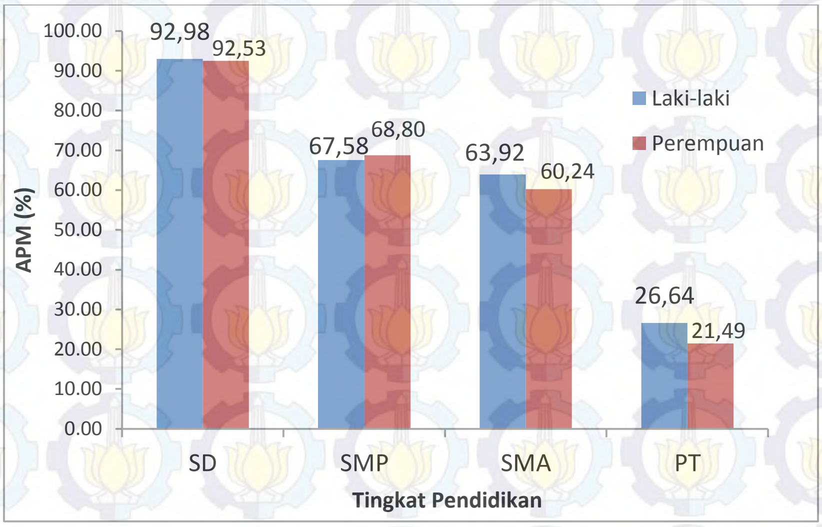
4.1. Gambaran Umum

Provinsi Papua Barat merupakan provinsi termuda keempat di Indonesia yang juga pemekaran dari Provinsi Papua pada tahun 2001. Pada tahun 2013, wilayah Provinsi Papua Barat terbagi dalam wilayah administrasi yang terdiri atas 10 (sepuluh) kabupaten dan 1 (satu) kota, 162 kecamatan, 1.322 desa serta 70 kelurahan. Luas wilayah Provinsi Papua Barat mencapai 97.407,61 km^2 (berdasarkan Peraturan Menteri Dalam Negeri Nomor 6 Tahun 2008). Jumlah penduduk Papua Barat yaitu sebesar 828.293 jiwa dengan pertumbuhan penduduk sebesar 2,89%.

Capaian pembangunan pada sektor pendidikan di Provinsi Papua Barat dapat diukur dari besarnya angka partisipasi murni dan angka partisipasi sekolah. Pada sub bab berikut akan dijelaskan tentang angka partisipasi murni dan angka partisipasi sekolah penduduk usia 7 – 24 tahun.

4.1.1. Angka Partisipasi Murni

Dalam publikasi Indeks Kesejahteraan Rakyat Provinsi Papua Barat, partisipasi sekolah dari penduduk usia sekolah sesuai dengan jenjang pendidikannya yang diukur dari Angka Partisipasi Murni (APM). Misalnya APM SD mengukur partisipasi sekolah penduduk usia 7-12 tahun yang masih bersekolah SD/sederajat, APM SMP mengukur partisipasi sekolah penduduk usia 13-15 tahun yang masih bersekolah SMP/sederajat, dan seterusnya. Gambar 4.1 berikut menunjukkan perbedaan APM antara anak laki-laki dan perempuan di jenjang pendidikan dasar, menengah, dan pendidikan tinggi.

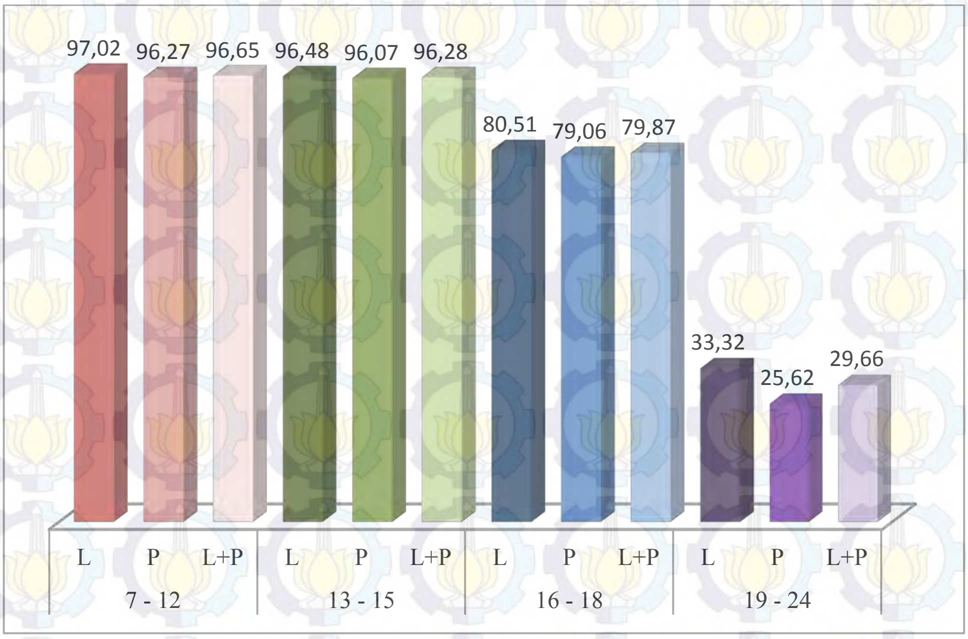


Gambar 4.1 Angka Partisipasi Murni Penduduk Menurut Jenis Kelamin di Provinsi Papua Barat

Dari histogram di atas, tampak bahwa pada usia sekolah dasar sebagian besar anak usia 7-12 tahun statusnya masih bersekolah. Selanjutnya pada tingkat di atasnya, yakni SMP, hanya 67,58 persen anak usia 13-15 tahun yang masih bersekolah untuk laki-laki dan 68,80 persen untuk perempuan. Artinya masih ada sekitar 30 persen anak pada usia tersebut yang belum bersekolah. Pada usia anak sekolah setingkat SMA/SMK, persentase APM juga menurun sebesar 63,92 persen untuk anak laki-laki dan 60,24 untuk anak perempuan. Persentase APM ini terus menurun hingga jenjang pendidikan tinggi. Ini berarti bahwa capaian APM di Provinsi Papua Barat bisa dikatakan belum cukup baik dan perlu ditingkatkan. Kaitannya dalam penelitian ini, rentang usia yang diamati dalam pemodelan angka lama sekolah di Provinsi Papua Barat yaitu anak sekolah usia 16-24 tahun untuk melihat seberapa efektif program Wajar 9 Tahun yang telah berjalan mengingat masih rendahnya persentase anak usia sekolah yang masih bersekolah di jenjang SMA dan pendidikan tinggi.

4.1.2. Angka Partisipasi Sekolah

Angka partisipasi sekolah (APS) mengukur proporsi dari semua anak yang masih sekolah pada satu kelompok umur tertentu terhadap penduduk dengan kelompok umur yang sesuai. Jadi indikator ini berguna untuk menunjukkan besarnya tingkat partisipasi pendidikan menurut kelompok umur tertentu, misalkan kelompok umur 7-12 tahun, 13-15 tahun, 16-18 tahun hingga 19-24 tahun.



Gambar 4.2 Angka Partisipasi Sekolah Penduduk 7 – 24 Tahun Menurut Jenis Kelamin di Provinsi Papua Barat

Berdasarkan Gambar 4.2 diperoleh informasi bahwa pada tahun 2014, 79,87 persen penduduk usia 16-18 tahun berstatus masih sekolah, sedangkan penduduk usia 19-24 tahun sebanyak 29,66 persen yang masih berstatus masih bersekolah. Perbedaan APS pada kedua kelompok usia ini sangat dipengaruhi oleh jumlah SMA/SMK serta Perguruan Tinggi (PT). Fasilitas pendidikan SMA/SMK/PT banyak terpusat di daerah perkotaan khususnya Kabupaten Manokwari dan Kota Sorong.

Baik pada kelompok usia 16-18 maupun 19-24 tahun keduanya masih didominasi oleh penduduk laki-laki yang lebih besar persentase partisipasi sekolahnya. Pada kelompok usia anak-anak 7-12 tahun dan kelompok usia SMP yaitu 13-15 tahun hampir tidak ada perbedaan partisipasi sekolah antara laki-laki dan perempuan. Namun pada kelompok usia 19-24 tahun tampak perbedaan yang cukup nyata antara partisipasi sekolah laki-laki dengan perempuan. Keterbatasan jumlah SMA dan PT berdampak pada partisipasi sekolah penduduk usia 16-24 tahun khususnya pada penduduk perempuan.

4.1.3. Karakteristik Penduduk Usia Sekolah 16 – 24 Tahun di Papua Barat

Untuk mengetahui karakteristik penduduk usia sekolah 16 - 24 tahun di Provinsi Papua Barat dapat dilakukan melalui analisis deskriptif terhadap masing-masing variabel yang digunakan dalam penelitian ini. Deskripsi untuk angka lama sekolah penduduk usia 16-24 tahun di Provinsi Papua Barat berdasarkan variabel-variabel prediktor yang berskala rasio menurut daerah perkotaan dan perdesaan seperti ditunjukkan tampak seperti pada Tabel 4.1 dan 4.2 berikut.

Tabel 4.1 Deskriptif Penduduk Usia 16 – 24 Tahun Daerah Perkotaan

Variabel	Minimum	Maksimum	Mean	Std. Deviasi
Lama Sekolah Penduduk Usia 16 - 24 tahun	0	16	11,08	2,71
Jumlah ART	2	18	5,89	2,49
Rata-rata Pengeluaran Perkapita Sebulan (dlm jutaan rupiah)	0,2594	5,9998	1,0637	0,7133

Berdasarkan Tabel 4.1 didapatkan informasi bahwa waktu *survival* penduduk atau angka lama sekolah penduduk umur 16 - 24 tahun di daerah perkotaan paling sedikit 0 tahun, yang artinya belum pernah memperoleh pendidikan formal, sedangkan angka lama sekolah paling tinggi yaitu 16 tahun yang artinya sudah mendapatkan ijazah pada jenjang SMA. Rata-rata lama sekolah penduduk berusia 16 - 24 tahun sebesar 11,08 yang artinya sebagian besar penduduk pada rentang usia tersebut memiliki rata-rata lama sekolah di atas 9 tahun atau dikatakan sudah lulus pendidikan pada jenjang SMP/sederajat dan sudah mencapai pendidikan setingkat kelas 2 SMA.

Sementara itu pada variabel jumlah anggota rumah tangga (ART), paling sedikit ada rumah tangga yang memiliki 2 ART dan yang paling banyak memiliki ART sebanyak 18 orang. Hal ini bisa dikatakan perbedaannya sangat ekstrim. Begitu pula pada variabel rata-rata pengeluaran perkapita perbulan, paling kecil sebesar Rp 259.400,- sedangkan paling besar rata-rata pengeluaran perkapita perbulan hingga mencapai Rp 5.999.800,-.

Tabel 4.2 Deskriptif Penduduk Usia 16 – 24 Tahun Daerah Perdesaan

Variabel	Minimum	Maksimum	Mean	Std. Deviasi
Lama Sekolah Penduduk Usia 16 - 24 tahun	0	16	9,43	3,42
Jumlah ART	2	20	5,92	2,47
Rata-rata Pengeluaran Perkapita Sebulan (dlm jutaan rupiah)	0,1448	8,7548	0,5572	0,4638

Di daerah perdesaan, didapatkan informasi bahwa waktu *survival* penduduk atau angka lama sekolah penduduk umur 16 - 24 tahun minimal 0 tahun yang artinya belum pernah memperoleh pendidikan formal sedangkan maksimal yaitu 16 tahun yang artinya sudah mendapatkan ijazah pada jenjang SMA. Sama halnya dengan di perkotaan, penduduk berusia 16 - 24 tahun di perdesaan sebagian besar dapat dikatakan sudah lulus pendidikan pada jenjang SMP/sederajat dimana rata-rata lama sekolah pada rentang usia tersebut sebesar 9,43.

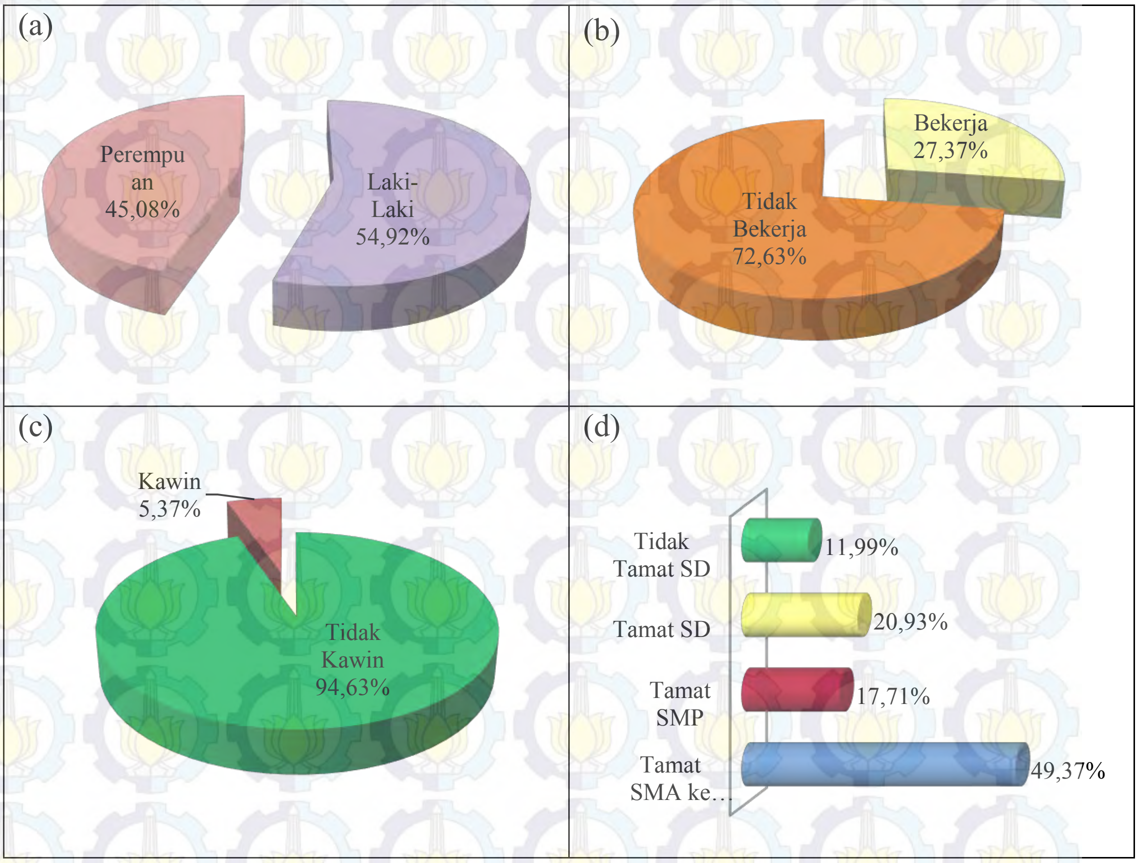
Sementara itu variabel jumlah anggota rumah tangga (ART), paling sedikit ada rumah tangga yang memiliki 2 ART dan yang paling banyak memiliki ART sebanyak 20 orang. Hal ini bisa dikatakan perbedaannya sangat ekstrim. Begitu pula pada variabel rata-rata pengeluaran perkapita perbulan, paling kecil sebesar Rp 144.800,- sedangkan paling besar rata-rata pengeluaran perkapita perbulan hingga mencapai Rp 8.754.800,-.

Untuk variabel prediktor yang bersifat kategorik, rata-rata lama sekolah berdasarkan masing-masing kategori menurut daerah perkotaan dan perdesaan dapat ditunjukkan pada Tabel 4.3 dan 4.4.

Tabel 4.3 Rata-Rata Lama Sekolah Daerah Perkotaan

Variabel Prediktor	Kategori	Rata-rata Lama Sekolah
Jenis Kelamin	Laki-laki	10,92
	Perempuan	11,29
Status Bekerja	Bekerja	11,38
	Tidak Bekerja	10,97
Status Perkawinan	Tidak Kawin	11,12
	Kawin	10,37
Pendidikan Kepala Rumah Tangga	Tidak Tamat SD	10,03
	Tamat SD	10,49
	Tamat SMP	10,79
	Tamat ≥ SMA	11,70

Tabel 4.3 menunjukkan besarnya rata-rata lama sekolah di daerah perkotaan apabila dirinci berdasarkan kategori tiap-tiap variabel prediktor. Rata-rata lama sekolah untuk usia penduduk 16 - 24 tahun di perkotaan tampak bahwa perempuan masih lebih dominan dibandingkan laki-laki. Sementara rata-rata lama sekolah penduduk usia 16-24 yang berstatus sudah bekerja memiliki angka yang sedikit lebih tinggi dibandingkan yang tidak bekerja. Penduduk usia 16 – 24 tahun yang tidak kawin memiliki rata-rata lama sekolah yang sedikit lebih tinggi dibandingkan yang sudah kawin. Dari sisi latar belakang pendidikan kepala rumah tangga, baik penduduk usia 16 – 24 tahun di daerah perkotaan dengan pendidikan kepala rumah tangga tidak tamat SD memiliki rata-rata lama sekolah yang hampir sama dibandingkan dengan penduduk yang memiliki kepala rumah tangga dengan tingkat pendidikan manapun. Pada tingkat pendidikan kepala rumah tangga manapun keempat kategori tersebut telah mampu memenuhi target pendidikan dasar 9 tahun.



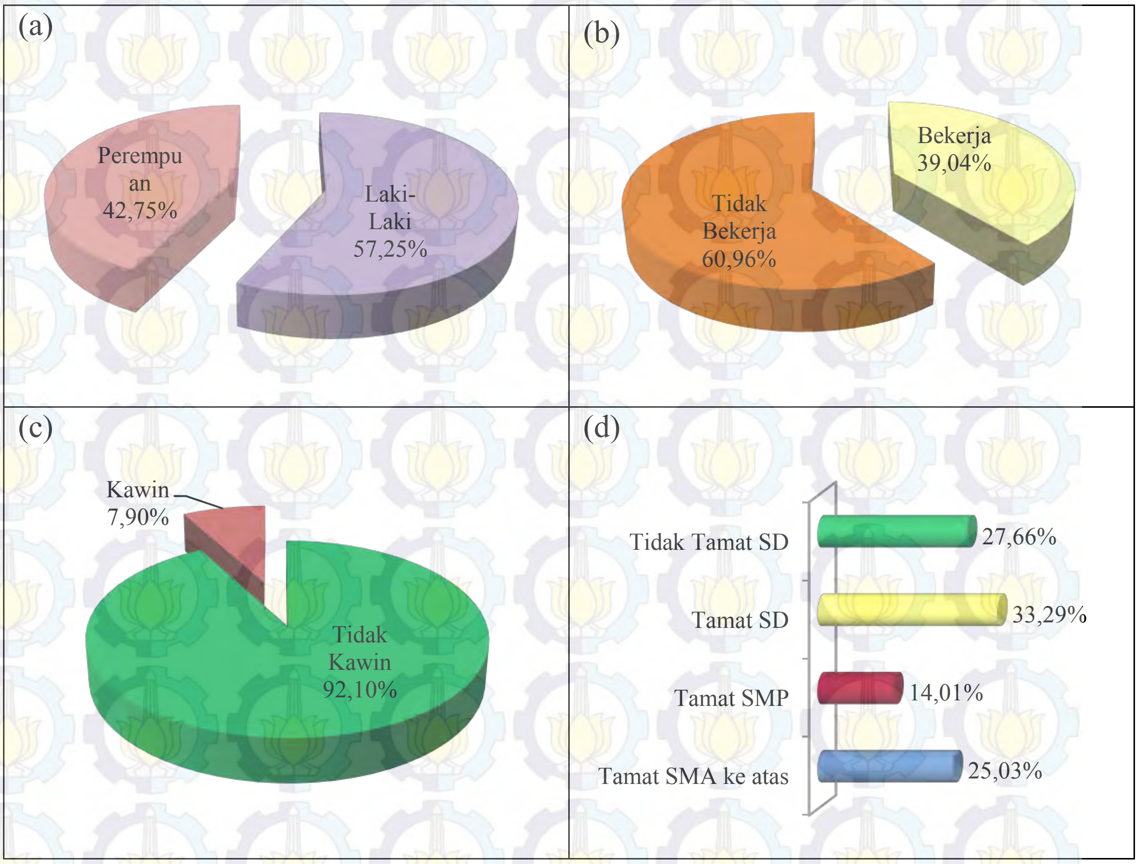
Gambar 4.3. Persentase Penduduk Usia 16 – 24 Tahun di Daerah Perkotaan berdasarkan (a) Jenis Kelamin, (b) Status Bekerja, (c) Status Perkawinan, dan (d) Pendidikan Kepala Rumah Tangga

Di daerah perkotaan, penduduk usia 16-24 tahun lebih banyak didominasi oleh penduduk laki-laki yaitu mencapai 54,92 persen, sedangkan sisanya adalah penduduk perempuan. Dan pada rentang usia tersebut sekitar 72,63 persen penduduk masih berstatus tidak bekerja. Hal ini wajar karena pada rentang usia tersebut merupakan masa-masa sekolah SMA ataupun masa kuliah baik D1, D3 maupun S1. Sementara itu, di daerah perkotaan masih sangat sedikit penduduk dalam rentang usia tersebut yang berstatus kawin. Sekitar 94,63 persen penduduk usia 16 – 24 tahun adalah penduduk yang masih belum kawin, bisa jadi statusnya cerai namun kemungkinannya sangatlah kecil. Jika ditinjau dari pendidikan kepala rumah tangga penduduk usia tersebut, hampir 50 persen di perkotaan latar pendidikan kepala rumah tangganya adalah SMA ataupun lebih, yakni meliputi D1, D2, D3, S1, S2 maupun S3. Hal ini sangat dimaklumi mengingat fasilitas pendidikan di perkotaan jauh lebih lengkap dibandingkan di perdesaan, seperti tersedianya gedung SMA, kampus/universitas khususnya di wilayah Kabupaten Manokwari dan Kota Sorong. Sedangkan yang kepala rumah tangganya belum pernah sekolah atau tidak menamatkan pendidikan SD hanya sekitar 11,99 persen.

Tabel 4.4 Rata-Rata Lama Sekolah Daerah Perdesaan

Variabel Prediktor	Kategori	Rata-rata Lama Sekolah
Jenis Kelamin	Laki-laki	9,33
	Perempuan	9,57
Status Bekerja	Bekerja	8,52
	Tidak Bekerja	10,01
Status Perkawinan	Tidak Kawin	9,44
	Kawin	9,33
Pendidikan Kepala Rumah Tangga	Tidak Tamat SD	7,66
	Tamat SD	9,31
	Tamat SMP	10,32
	Tamat ≥ SMA	11,04

Jika pada Tabel 4.3 menunjukkan rata-rata lama sekolah di daerah perkotaan apabila dirinci berdasarkan kategori tiap-tiap variabel prediktor, maka Tabel 4.4 menunjukkan besarnya rata-rata lama sekolah di daerah perdesaan. Sama halnya dengan di perkotaan, rata-rata lama sekolah penduduk usia 16 - 24 tahun di perdesaan tampak bahwa perempuan sedikit lebih lama dibandingkan laki-laki namun tidak tampak perbedaan yang cukup berarti. Lain halnya dengan di perkotaan, justru penduduk usia 16 - 24 di perdesaan yang berstatus tidak bekerja memiliki rata-rata angka lama sekolah yang lebih tinggi dibandingkan yang sedang bekerja. Sementara itu, penduduk pada rentang usia tersebut yang tidak kawin dengan yang kawin memiliki rata-rata lama sekolah yang hampir sama walaupun agak sedikit lebih lama penduduk yang tidak kawin. Dari sisi latar belakang pendidikan kepala rumah tangga, tampak bahwa penduduk usia 16 – 24 tahun di daerah perdesaan dengan pendidikan kepala rumah tangga tidak tamat SD memiliki rata-rata lama sekolah yang sangat kecil dibandingkan dengan penduduk yang memiliki kepala rumah tangga dengan tingkat pendidikan lebih tinggi. Pada daerah perdesaan ini semakin jelas bahwa terdapat hubungan yang positif antara peningkatan pendidikan kepala rumah tangga dengan rata-rata lama sekolah yang berhasil ditempuh oleh anaknya.



Gambar 4.4. Persentase Penduduk Usia 16 – 24 Tahun di Daerah Perdesaan berdasarkan (a) Jenis Kelamin, (b) Status Bekerja, (c) Status Perkawinan, dan (d) Pendidikan Kepala Rumah Tangga

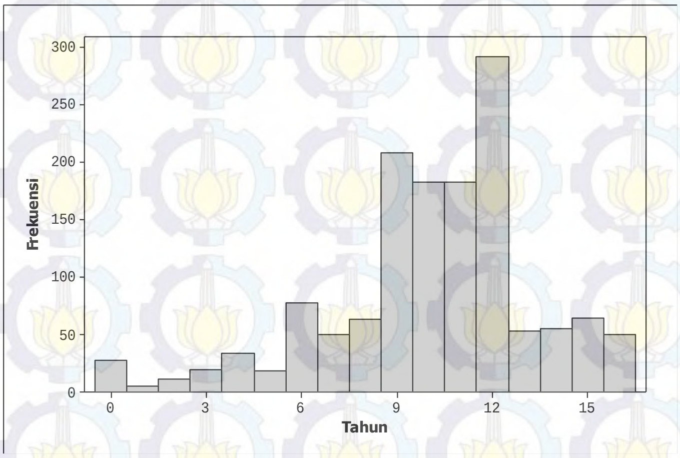
Sama halnya dengan di perkotaan, penduduk usia 16-24 tahun di daerah perdesaan lebih banyak didominasi oleh penduduk laki-laki yaitu mencapai 57,25 persen, sedangkan sisanya adalah penduduk perempuan. Dan pada rentang usia tersebut sekitar 60,96 persen penduduk masih berstatus tidak bekerja. Hal ini wajar karena pada rentang usia tersebut merupakan masa-masa sekolah SMA ataupun masa kuliah baik D1, D3 maupun S1. Akan tetapi jika dibandingkan dengan penduduk pada rentang usia yang sama di perkotaan, persentase penduduk yang bekerja masih lebih tinggi di perdesaan. Hal ini dapat dimaklumi mengingat di daerah perdesaan banyak penduduk pada usia tersebut yang membantu pekerjaan orang tuanya yang dominan adalah pekerja di bidang pertanian baik dari subsektor tanaman pangan maupun perikanan.

Sementara itu, di daerah perdesaan masih sangat sedikit penduduk dalam rentang usia tersebut yang berstatus kawin. Terdapat sekitar 92,10 persen penduduk usia 16 – 24 tahun adalah penduduk yang masih belum kawin, bisa jadi statusnya cerai namun kemungkinannya sangatlah kecil. Jika ditinjau dari pendidikan kepala rumah tangga penduduk usia tersebut, persentase tertinggi adalah kepala rumah tangga dengan latar belakang pendidikan yang diraih yaitu tamat SD (sekitar 33,29 persen). Hanya sekitar 25,03 persen kepala rumah tangga di perdesaan memiliki latar pendidikan setingkat SMA atau di atasnya.Hal ini bisa disebabkan karena jauhnya fasilitas pendidikan dari tempat tinggal mereka atau dengan kata lain minimnya fasilitas pendidikan di daerah perdesaan.Peningkatan yang cukup drastis antara kepala rumah tangga yang tamatan SMP dengan yang tamat pendidikan lebih tinggi yang mencapai dua kali lipatnya dimana pada tamatan SMP sebesar 14,01 persen sementara yang tamatan SMA ke atas hingga 25 persen.

4.2. Distribusi Data Lama Sekolah

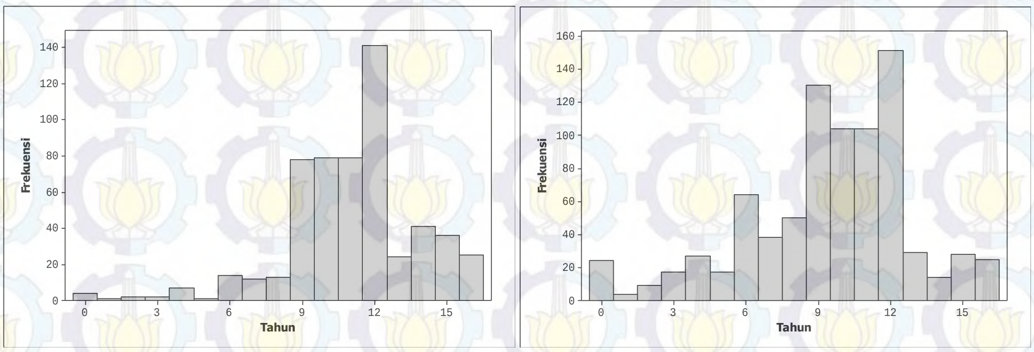
Provinsi Papua Barat yang terdiri atas 11 kabupaten/kota sebagian merupakan daerah pantai dan sisanya merupakan daerah pegunungan. Daerah-daerah pantai banyak didominasi oleh daerah kepulauan seperti Kabupaten Raja Ampat, Kabupaten Kaimana, dan Kabupaten Fakfak. Sedangkan daerah pegunungan banyak ditemukan di sebagian wilayah Kabupaten Manokwari, Kabupaten Tambrauw, Kabupaten Maybrat, Kabupaten Sorong Selatan.

Semantara itu, berdasarkan Perka BPS No 37 Tahun 2010 jumlah wilayah perkotaan di Papua Barat yaitu sebanyak 29 desa/kelurahan sedangkan wilayah perdesaan mencapai 1.338 desa/kelurahan. Berdasarkan data Susenas 2014, jumlah penduduk usia 16 – 24 tahun di Provinsi Papua Barat yang berstatus anak yaitu sebanyak 1.394 orang dengan 559 orang di daerah perkotaan dan sisanya di daerah perdesaan.



Gambar 4.5. Histogram Data Lama Sekolah di Provinsi Papua Barat

Gambar 4.5 menunjukkan pola distribusi data lama sekolah penduduk usia 16 – 24 tahun 2014 di Provinsi Papua Barat yang berstatus anak. Pada gambar tersebut tampak bahwa data cenderung tidak simetris, memiliki dua puncak yaitu di lama sekolah 9 tahun dan 12 tahun. Hal ini mengindikasikan adanya sifat multimodal pada data yang menyebabkan data tersebut berdistribusi *mixture*. Dalam penelitian ini komponen *mixture* diduga berdasarkan klasifikasi daerah tempat tinggal yaitu daerah perkotaan dan perdesaan dengan pola distribusi data seperti pada Gambar 4.6. Selanjutnya dalam tahap awal penelitian ini akan dilakukan pengujian distribusi data untuk mengetahui apakah data mengikuti pola distribusi tertentu terhadap seluruh wilayah Provinsi Papua Barat, perkotaan maupun perdesaan.



Gambar 4.6. Histogram Data Lama Sekolah di Daerah (a) Perkotaan dan (b) Perdesaan

Histogram data pada Gambar 4.6 menunjukkan pola data lama sekolah di daerah perkotaan maupun perdesaan. Selanjutnya Tabel 4.5 menunjukkan nilai statistik uji Anderson-Darling pada pengujian distribusi data (*goodness of fit*) pada taraf signifikansi 5 % yang terkecil untuk data waktu *survival* lama sekolah secara keseluruhan, wilayah perkotaan maupun perdesaan

Tabel 4.5. *Goodness of Fit* Distribusi Lama Sekolah

Distribusi	Papua Barat	Perkotaan	Perdesaan	Nilai Kritis
Normal	25,536	9,461	14,012	2,5018
Weibull	**74,280**	**16,566**	**57,506**	**2,5018**
Lognormal	120,470	29,102	83,742	2,5018
Log-Logistik	97,523	18,360	75,025	2,5018
Gamma	99,292	21,998	71,358	2,5018

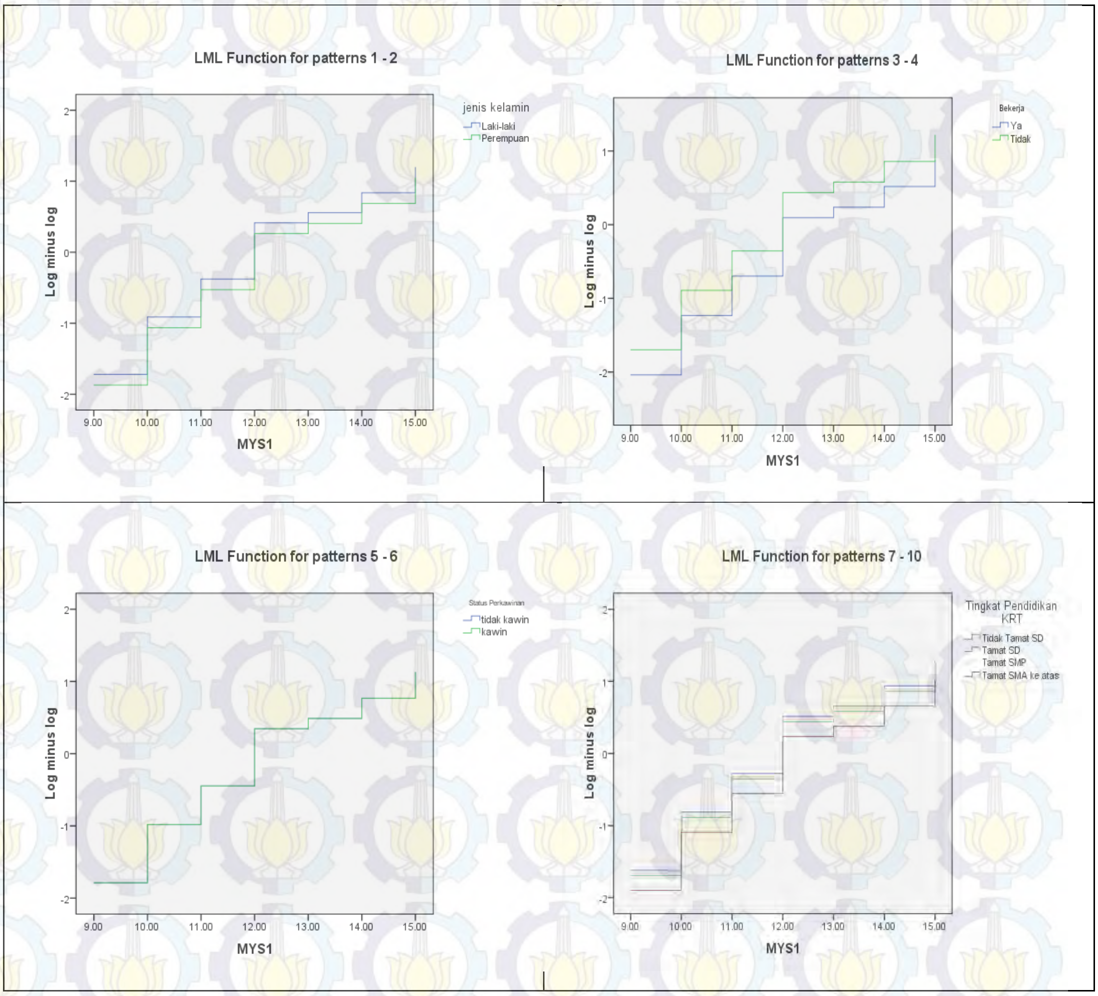
Hasil pengujian *goodness of fit* pada Tabel 4.5 tampak bahwa nilai statistik uji Anderson-Darling pada taraf signifikansi 5% diketahui bahwa tidak ada distribusi yang sesuai baik secara keseluruhan, maupun daerah perkotaan dan perdesaan karena nilai statistik uji Anderson Darling lebih besar dari nilai kritis. Pada pengujian tersebut nilai statistik uji Anderson Darling yang terkecil untuk data waktu *survival* lama sekolah secara keseluruhan, wilayah perkotaan maupun perdesaan adalah distribusi normal, yakni dengan nilai statistik uji masing-masing sebesar 25,536; 9,461 dan 14,012. Namun, dalam penelitian ini, pola distribusi *weibull* lebih dipertimbangkan karena pola distribusi *weibull* lebih sesuai dipergunakan pada data *survival* dibandingkan distribusi normal yang memiliki nilai distribusi berkisar antara -∞ sampai ∞. Adanya indikasi distribusi *mixture* dapat dipertimbangkan dimana pendekatan yang digunakan adalah distribusi *weibull* untuk kedua wilayah.

4.3. Model *Survival*

Tahapan yang dilakukan setelah diketahui distribusi data lama sekolah yaitu pembentukan model regresi *survival* di daerah perkotaan dan perdesaan. Langkah-langkah yang dilakukan dalam pemodelan regresi *Cox proportional hazard* adalah sebagai berikut:

a. Memeriksa asumsi *proportional hazard* pada setiap variabel prediktor. Analisis regresi *Cox proportional hazard* ini tidak membutuhkan asumsi distribusi namun memerlukan pemenuhan asumsi *proportional hazard*, yaitu independensi variabel prediktor terhadap waktu lama sekolah. Asumsi ini dapat terpenuhi dengan melihat plot antara ln(-ln $S(t)$) terhadap t untuk setiap variabel prediktor. Apabila garis antar kategori sejajar maka asumsi *proportional hazard* dapat dikatakan terpenuhi (Collet, 1994).

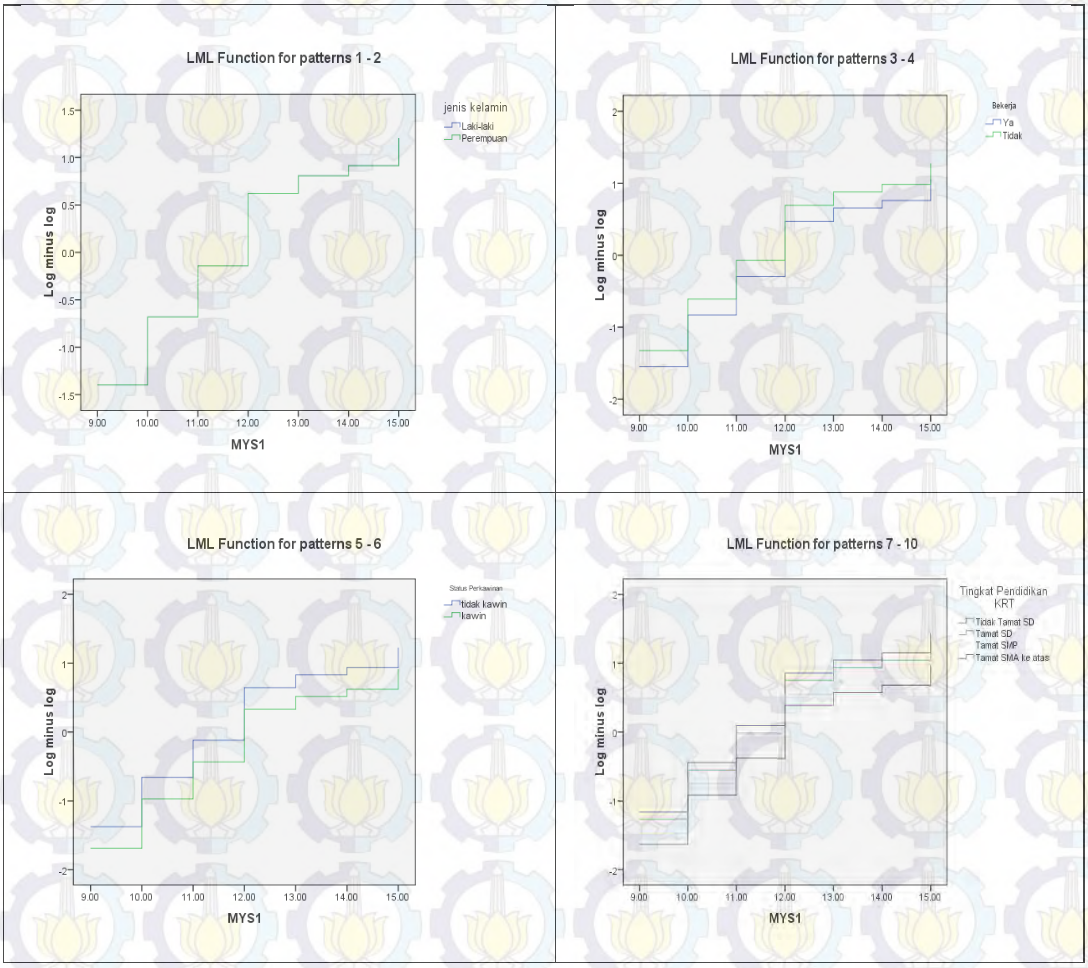
   Pada Gambar 4.7 dan 4.8 terlihat bahwa plot ln(-ln $S(t)$) terhadap lama sekolah pada masing-masing variabel prediktor kategorik untuk daerah perkotaan dan perdesaan.



Gambar 4.7. Plot Fungsi ln(-ln $S(t)$) di Daerah Perkotaan

Berdasarkan plot fungsi ln(-ln $S(t)$) pada masing-masing variabel prediktor di daerah perkotaan, keempat variabel prediktor tersebut memenuhi asumsi *proportional hazard*, yaitu variabel jenis kelamin, status bekerja, status perkawinan, dan tingkat pendidikan kepala rumah tangga. Variabel-variabel tersebut memiliki pola yang sejajar antar kategori pada masing-masing variabel. Artinya variabel-variabel prediktor tersebut telah memenuhi asumsi *proportional hazard*.

Selanjutnya pengujian asumsi *proportional hazard* untuk daerah perdesaan dapat ditunjukkan pada Gambar 4.8 berikut.



Gambar 4.8. Plot Fungsi ln(-ln $S(t)$) di Daerah Perdesaan

Plot fungsi ln(- ln $S(t)$) pada keempat variabel prediktor kategorik untuk daerah perdesaan memiliki pola yang sejajar dan tidak berpotongan antar kategori. Hal ini dapat disimpulkan bahwa varaibel prediktor tersebut telah memenuhi asumsi *proportional hazard*.

b. Melakukan pemilihan model *Cox proportional hazard* terbaik. Pemilihan model terbaik menurut Collet (1994) dapat dilakukan dengan cara metode eliminasi *backward* dengan dasar uji *likelihood ratio*. Tahapan-tahapan yang dilakukan dengan metode eliminasi *backward* yaitu:
   (i) Menyusun model yang terdiri atas seluruh variabel prediktor yang memenuhi asumsi *proportional hazard* secara bersamaan.
   (ii) Pada setiap langkah, variabel yang dikeluarkan adalah variabel dengan nilai yang berasal dari perubahan -2 ln $\hat{L}$ dari model sebelumnya.
   (iii) Proses akan berhenti ketika model selanjutnya memiliki peningkatan nilai -2 ln $\hat{L}$ melebihi dari model sebelumnya.

   Model regresi *cox* angka lama sekolah di Provinsi Papua Barat daerah perkotaan dengan melibatkan seluruh variabel prediktor adalah sebagai berikut.

$$\hat{h}_i(t) \propto \exp(0{,}163\,X_{1i} - 0{,}360\,X_{2i} - 0{,}276\,X_{3i} + 0{,}235\,X_{41i} + 0{,}217\,X_{42i} + 0{,}245\,X_{43i} + 0{,}011\,X_{5i} - 0{,}001\,X_{6i})$$

   Model regresi *Cox proportional hazard* terbaik diperoleh dengan metode eliminasi *backward* berdasarkan rasio *likelihood*-nya. Proses tersebut secara rinci dapat dilihat pada Tabel 4.6.

Tabel 4.6 Pemilihan Model Terbaik Daerah Perkotaan

Step	**Variabel dalam Model**	**-2Ln $\hat{L}$**	**Uji Serentak**			**Perubahan setiap Tahap**		
			χ^2	**df**	***p-value***	**χ^2**	**df**	***p-value***
1	*null*	5494,300						
2	$X_1,X_2,X_3,X_{41},X_{42},X_{43},X_5,X_6$	5474,594	19,306	8	0,013	19,706	8	0,012
3	$X_1,X_2,X_3,X_{41},X_{42},X_{43},X_5$	5474,616	19,288	7	0,007	0,022	1	0,882
4	$X_1,X_2,X_3,X_{41},X_{42},X_{43}$	5474,056	18,750	6	0,005	0,440	1	0,507
5	$X_1,X_2,X_{41},X_{42},X_{43}$	5474,938	16,926	5	0,005	1,882	1	0,170

Keterangan:
X_1 : jenis kelamin
X_2 : status bekerja
X_3 : status perkawinan
X_{41} : pendidikan KRT tidak tamat SD
X_{42} : pendidikan KRT tamat SD
X_{43} : pendidikan KRT tamat SMP
X_5 : jumlah ART
X_6 : rata-rata pengeluaran perkapita 1 bln

Hasil pengujian serentak pada Tabel 4.6 diperoleh variabel-variabel yang masuk dalam model yang berpengaruh terhadap angka lama sekolah untuk daerah perkotaan di Provinsi Papua Barat. Terlihat bahwa model pada step 1 merupakan *null* model dimana model tersebut tidak memuat variabel prediktor. Selanjutnya dilakukan langkah awal yaitu memasukkan seluruh variabel prediktor ke dalam model. Dari langkah tersebut maka didapatkan nilai -2ln $\hat{L}$ sebesar 5474,594 dimana terlihat adanya perubahan nilai -2ln $\hat{L}$ dari *null* model sehingga terjadi perubahan nilai χ^2 menjadi 19,306 dengan *p-value* sebesar 0,013. Dengan diperolehnya *p-value* tersebut, dibandingkan dengan α sebesar 5 persen maka dapat disimpulkan bahwa model signifikan. Selanjutnya model pada step 3 yaitu dengan mengeluarkan variabel X_6 (rata-rata pengeluaran perkapita sebulan) dari model karena variabel X_6 memberikan nilai χ^2 terkecil yaitu 0,022 dengan *p-value* sebesar 0,882 > 5 persen. Proses tersebut terus berlanjut dengan mengeluarkan variabel prediktor dengan nilai χ^2 terkecil hingga diperoleh model dengan melibatkan variabel X_1, X_2, X_{41}, X_{42}, dan X_{43} dengan nilai χ^2 sebesar 16,926 dan *p-value* sebesar 0,005.

Sama halnya dengan di daerah perkotaan, di daerah perdesaan juga dibentuk model regresi *cox* angka lama sekolah dengan keseluruhan variabel prediktor sehingga didapatkan model sebagai berikut:

$$\begin{aligned}\hat{h}_i(t) \propto\ & \exp\,(0,045\,X_{1i} - 0,228\,X_{2i} + 0,294\,X_{3i} + 0,428\,X_{41i} + 0,327\,X_{42i} \\ & + 0,120\,X_{43i} - 0,005\,X_{5i} - 0,016\,X_{6i})\end{aligned}$$

Selanjutnya model regresi *Cox proportional hazard* terbaik diperoleh dengan metode eliminasi *backward* berdasarkan rasio *likelihood*-nya. Proses tersebut secara rinci dapat dilihat pada Tabel 4.7.

Tabel 4.7 Pemilihan Model Terbaik Daerah Perdesaan

Step	Variabel dalam Model	-2Ln $\hat{L}$	Uji Serentak			Perubahan setiap Tahap		
			χ^2	df	*p-value*	χ^2	df	*p-value*
1	*null*	6578,263						
2	$X_1,X_2,X_3,X_{41}, X_{42}, X_{43},X_5,X_6$	6551,539	26,080	8	0,001	26,724	8	0,001
3	$X_1,X_2,X_3,X_{41},X_{42},X_{43},X_6$	6551,593	26,038	7	0,000	0,054	1	0,817
4	$X_2,X_3,X_{41},X_{42},X_{43},X_6$	6551,899	25,745	6	0,000	0,305	1	0,581
5	$X_2,X_3,X_{41},X_{42},X_{43}$	6553,920	23,955	5	0,000	2,022	1	0,155

Keterangan:
X_1 : jenis kelamin
X_2 : status bekerja
X_3 : status perkawinan
X_{41} : pendidikan KRT tidak tamat SD
X_{42} : pendidikan KRT tamat SD
X_{43} : pendidikan KRT tamat SMP
X_5 : jumlah ART
X_6 : rata-rata pengeluaran perkapita 1 bln

Hasil pengujian serentak pada Tabel 4.7 diperoleh variabel-variabel yang masuk dalam model yang berpengaruh terhadap angka lama sekolah untuk daerah perdesaan di Provinsi Papua Barat. Pada Tabel 4.7 terlihat bahwa model pada step 1 merupakan *null* model dimana model tersebut tidak memuat variabel prediktor. Selanjutnya dilakukan langkah awal yaitu memasukkan seluruh variabel prediktor ke dalam model. Dari langkah tersebut didapatkan nilai -2ln $\hat{L}$ sebesar 6551,539 dimana terlihat adanya perubahan nilai -2ln $\hat{L}$ dari *null* model sehingga terjadi perubahan nilai χ^2 menjadi 26,080 dengan *p-value* sebesar 0,001. Dengan diperolehnya *p-value* tersebut, dan dibandingkan dengan α sebesar 5 persen maka dapat disimpulkan bahwa model signifikan. Selanjutnya model pada step 3 yaitu dengan mengeluarkan variabel X_5 (jumlah ART) dari model karena variabel X_5 memberikan nilai χ^2 terkecil yaitu 0,054 dengan *p-value* sebesar 0,817, yaitu lebih dari 5 persen. Proses tersebut terus berlanjut dengan mengeluarkan variabel prediktor dengan nilai χ^2 terkecil hingga diperoleh model dengan melibatkan variabel X_2, X_3, X_{41}, X_{42}, dan X_{43} dengan nilai χ^2 sebesar 23,955 dan *p-value* sebesar 0,000.

c. Melakukan uji parsial untuk masing-masing parameter.

Setelah dilakukan pengujian serentak untuk masing-masing daerah, maka selanjutnya yaitu melakukan pengujian parsial untuk masing-masing parameter pada daerah perkotaan dan perdesaan. Hasil pengujian untuk daerah perkotaan dapat diringkas seperti tabel dibawah ini.

Tabel 4.8. Koefisien Variabel yang Masuk dalam Model berdasarkan Metode Eliminasi *Backward* di Daerah Perkotaan

Variabel	**Koefisien** $(\hat{\beta})$	χ^2w	***Odd Ratio*** $(e^{\hat{\beta}})$	***p-value***
X_1	0,150	2,751	1,162	0,097
X_2	-0,340	10,885	0,711	0,001
X_{41}	0,278	3,226	1,321	0,072
X_{42}	0,205	2,942	1,227	0,086
X_{43}	0,238	3,752	1,269	0,053

Pada Tabel 4.8 terlihat bahwa variabel jenis kelamin (X_1), status bekerja (X_2), pendidikan KRT tidak tamat SD (X_{41}), pendidikan KRT tamat SD (X_{42}), dan pendidikan KRT tamat SMP (X_{43}) signifikan pada taraf 10 persen, sehingga model terbaik regresi *Cox proportional hazard*-nya adalah:

$$\hat{h}_i(t) \propto \exp\ (0,150X_{1i} - 0,340X_{2i} + 0,278X_{41i} + 0,205X_{42i} + 0,238X_{43i}).$$

Serupa dengan langkah yang dilakukan pada daerah perkotaan, hasil pengujian parsial untuk daerah perdesaan seperti ditampilkan pada Tabel 4.9.

Tabel 4.9. Koefisien Variabel yang Masuk dalam Model berdasarkan Metode Eliminasi *Backward* di Daerah Perdesaan

Variabel	**Koefisien** $(\hat{\beta})$	χ^2w	***Odd Ratio*** $(e^{\hat{\beta}})$	***p-value***
X_2	-0,223	5,777	0,800	0,016
X_3	0,314	3,692	1,368	0,055
X_{41}	0,471	14,334	1,602	0,000
X_{42}	0,363	11,374	1,437	0,001
X_{43}	0,163	1,617	1,178	0,204

Pada Tabel 4.9 terlihat bahwa variabel status bekerja (X_2), status perkawinan (X_3), pendidikan KRT tidak tamat SD (X_{41}), pendidikan KRT tamat SD (X_{42}), dan pendidikan KRT tamat SMP (X_{43}) signifikan pada taraf 10 persen, sehingga model terbaik regresi *Cox proportional hazard*-nya adalah:

$$\hat{h}_i\left(t\right) \propto \exp\ (-0,223X_{2i} + 0,314X_{3i} + 0,471X_{41i} + 0,363X_{42i} + 0,163X_{43i}).$$

4.4. Model *Mixture Survival*

Model *mixture* regresi *survival* dengan dua komponen didasarkan pada persamaan (2.28) dimana fungsi densitasnya disusun dari distribusi data *survival* (*t*) sehingga diperoleh bentuk seperti pada persamaan (2.34) dapat juga dinyatakan sebagai berikut:

$$p\left(t\,|\,\pi,\theta\right) = \pi\, p\left(t|\theta_1\right) + (1-\pi)\, p\left(t|\theta_2\right),$$

dimana pada tahap awal pembentukan *mixture survival* adalah dengan melakukan uji distribusi data (*goodness of fit*) data *survival*. Persamaan fungsi *survival* untuk distribusi *mixture* dengan dua subpopulasi (komponen) dapat ditulis sebagai berikut:

$$S(t) = \pi\, S_1(t) + (1-\pi)\, S_2(t), \tag{4.1}$$

sehingga model *proportional hazards* untuk *mixture survival* dapat dinyatakan sebagai berikut:

$$h(t) = \pi\, h_1(t) + (1-\pi)\, h_2(t), \tag{4.2}$$

dengan $h_1(t)$ adalah model *mixture proportional hazards* subpopulasi pertama (perkotaan) dan $h_2(t)$ merupakan model *mixture proportional hazards* subpopulasi kedua (perdesaan).

Densitas distribusi *mixture weibull* sesuai dengan persamaan (2.34) menjadi:

$$p(t\,|\,\pi,\theta) = \pi((\lambda_1\gamma_1 t^{\gamma_1 - 1}\exp(-\lambda_1 t^{\gamma_1})) + (1-\pi)(\lambda_2\gamma_2 t^{\gamma_2 - 1}\exp(-\lambda_2 t^{\gamma_2})). \tag{4.3}$$

Fungsi *survival* dalam model *mixture weibull* tersebut adalah:

$$S(t) = \pi\exp(-\lambda_1 t^{\gamma_1}) + (1-\pi)\exp(-\lambda_2 t^{\gamma_2}). \tag{4.4}$$

Umumnya, bentuk fungsi *hazards* berdasarkan distribusi *mixture weibull* dengan λ sebagai parameter skala dan γ sebagai parameter bentuk dapat ditulis sebagai berikut:

$$h(t) = \pi \lambda_1 \gamma_1 t^{\gamma_1 - 1} + (1-\pi)\lambda_2 \gamma_2 t^{\gamma_2 - 1} \tag{4.5}$$

untuk $0 \leq t \leq \infty$. Fungsi tersebut bergantung pada dua parameter yaitu λ dan γ untuk masing-masing subpopulasi dimana nilai $\lambda > 0$ dan $\gamma > 0$.

Nilai variabel prediktor dalam model *proportional hazards* yaitu x_1, x_2, …, x_p, sehingga model regresi *weibull* yang dapat dibentuk dari persamaan (2.8) untuk model tunggal menjadi seperti berikut:

$$h_i(t) = \lambda \gamma t^{\gamma - 1} \exp\left(\beta_1 x_{1i} + \ldots + \beta_p x_{pi}\right). \tag{4.6}$$

Pada model *proportional hazard* di (4.6), parameter yang berpengaruh adalah parameter skala (λ) saja, sedangkan parameter bentuk (γ) tidak berubah. Oleh karenanya parameterisasi model dilakukan pada parameter λ. Apabila kumpulan variabel prediktor pada model *proportional hazard* diwakili oleh vektor **x**, yaitu **x** = (x_1, x_2, …, x_p) maka model *mixture weibull proportional hazard* untuk wilayah perkotaan dan perdesaan berdasarkan persamaan (4.2) dan (4.6) menjadi:

$$h_i(t) = \pi \lambda_1 \gamma_1 t^{\gamma_1 - 1} \exp(\boldsymbol{\beta}_1^{'} \mathbf{x}_{i1}) + (1-\pi)\lambda_2 \gamma_2 t^{\gamma_2 - 1} \exp(\boldsymbol{\beta}_2^{'} \mathbf{x}_{i2}), \tag{4.7}$$

dan fungsi *survival* untuk model *mixture weibull proportional hazard* untuk wilayah perkotaan dan perdesaan sebagai berikut:

$$S^*(t) = \pi \exp(-\lambda_1 t^{\gamma_1}) \exp(\boldsymbol{\beta}_1^{'} \mathbf{x}_{i1}) + (1-\pi) \exp(-\lambda_2 t^{\gamma_2}) \exp(\boldsymbol{\beta}_2^{'} \mathbf{x}_{i2}). \tag{4.8}$$

4.4.1. Distribusi *Prior* dan *Joint* Distribusi *Posterior* dari Model *Mixture Survival*

Dalam model *mixture* setiap subpopulasi memiliki karakteristik data waktu *survival* masing-masing, dimana berdasarkan uji *goodness of fit* diperoleh distribusi data lama sekolah pada kedua wilayah (perkotaan dan perdesaan) memiliki distribusi *weibull* (γ,λ) dengan dua parameter yang berbeda untuk masing-masing wilayah.

Estimasi untuk setiap parameter diperoleh melalui metode *Bayesian* dengan menentukan distribusi *prior* terlebih dahulu. Distribusi *prior* yang digunakan dalam penelitian ini adalah *prior conjugate*, pseudo *prior*, dan *informative prior*.

Prior conjugate yakni *prior* yang terkait dengan pola *likelihood* dimana γ dan λ dari distribusi *weibull* adalah keluarga distribusi eksponensial. Distribusi gamma merupakan keluarga eksponensial dimana parameternya dapat berubah-ubah seperti halnya distribusi *weibull* sehingga digunakan distribusi gamma sebagai *prior* (Qian, 1994). Sedangkan distribusi *prior* untuk parameter *β* (dalam WinBUGS 1.4 dilambangkan dengan bi) menggunakan *prior* informatif yaitu distribusi normal. Nilai *prior* untuk parameter *β* merupakan pseudo *prior* yang didasarkan pada hasil pengolahan regresi *Cox proportional hazard* dengan bantuan *software* SPSS 16. Selanjutnya untuk distribusi *prior* untuk proporsi *mixture* menggunakan distribusi dirichlet.

Berdasarkan persamaan (2.34), dengan θ dijabarkan menjadi parameter distribusi *weibull* dua parameter (γ,λ), model *mixture* dengan penjabaran teori bayes distribusi posterior bersyarat penuh (*full conditional posterior*) akan menjadi proporsional terhadap fungsi *likelihood* dikali dengan *prior*. Dalam bentuk persamaan dapat dituliskan sebagai berikut:

$$p(\lambda,\gamma,\pi \mid t) \propto l(t \mid \lambda,\gamma,\pi)\times p(\lambda)\times p(\gamma)\times p(\pi) \tag{4.9}$$

dengan:

$$l(t \mid \boldsymbol{\lambda},\boldsymbol{\gamma},\boldsymbol{\pi}) = \pi\prod_{i_1=1}^{n_1}\lambda_1\gamma_1 t_{i_1}^{\gamma_1-1}\exp(-\lambda_1 t_{i_1}^{\gamma_1}) + (1-\pi)\prod_{i_2=1}^{n_2}\lambda_2\gamma_2 t_{i_2}^{\gamma_2-1}\exp(-\lambda_2 t_{i_2}^{\gamma_2}),$$

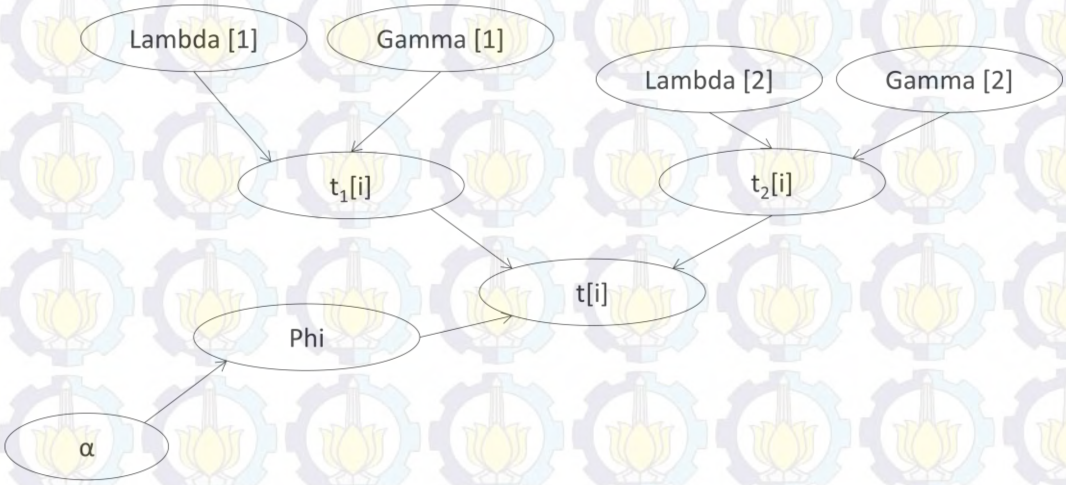
yang merupakan perkalian dari fungsi densitas distribusi *weibull* (γ, λ) dimana $t > 0$,

$p(\lambda) = \dfrac{1}{b^a\Gamma(a)}\lambda^{a-1}\exp(-\dfrac{\lambda}{b})$ adalah distribusi *prior* bagi γ, yaitu Gamma(α,b) dengan nilai $\lambda > 0$.

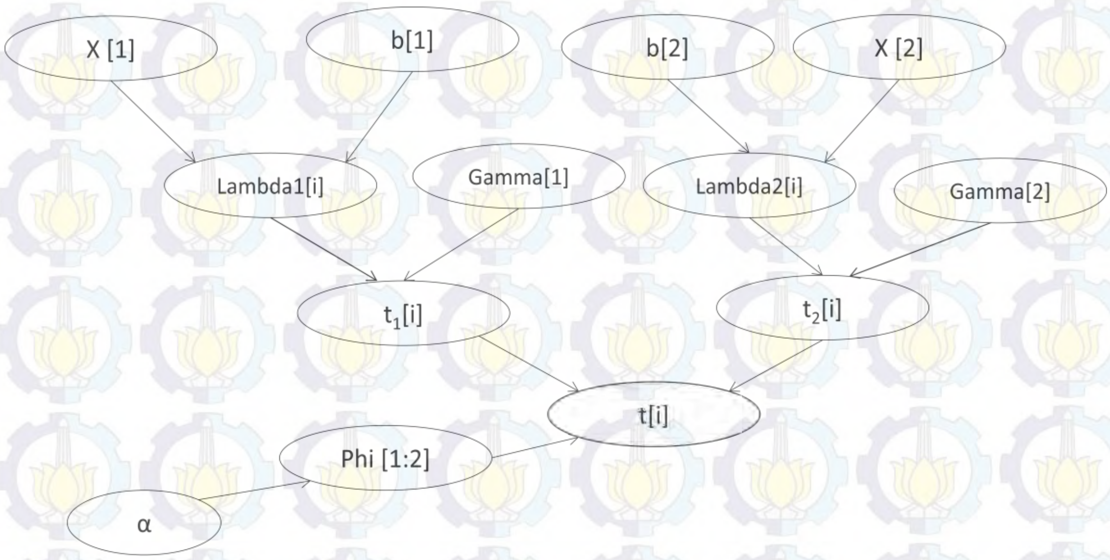
$p(\gamma) = \dfrac{1}{b^a\Gamma(a)}\gamma^{a-1}\exp(-\dfrac{\gamma}{b})$ adalah distribusi *prior* bagi γ, yaitu Gamma(α,b) dengan nilai $\gamma > 0$, dan

$p(\pi) = \frac{\Gamma\left(\sum_m \alpha_m\right)}{\prod_m \Gamma(\alpha_m)} \prod_m \pi_m^{\alpha_m - 1}$ adalah distribusi *prior* bagi proporsi *mixture* (π[ ]) yaitu Dirichlet (α[ ]) dimana $0 < \pi_m < 1$, $\sum_m \pi_m = 1$.

Berikut merupakan bentuk struktur parameter distribusi *mixture* dan model *mixture proportional hazard* yang akan diaplikasikan dengan *software* Winbugs 1.4.



Gambar 4.9. Struktur Parameter Distribusi *Mixture*



Gambar 4.10. Struktur Parameter Model *Mixture Proportional Hazard*

4.4.2. Estimasi Parameter Model *Mixture Survival* dengan menggunakan *Gibbs Sampling*

Estimasi parameter untuk distribusi *mixture* dilakukan dengan pendekatan distribusi *weibull* dengan menggunakan metode *Bayesian* melalui teknik MCMC *gibbs sampling*. Sebelum melakukan estimasi terhadap parameter π, λ, dan γ (dalam syntak secara berurutan disimbolkan dengan Phi, pLambda, dan pGamma), maka terlebih dahulu menentukan distribusi prior sebagai berikut,

π ~ Dirichlet (Alpha[ ])

$$\lambda \sim Gamma\ (a_m, b_m)$$

$$\gamma \sim Gamma\ \left(c_m, d_m\right)$$

dimana m = 1, 2 untuk banyaknya komponen penyusun *mixture*. Penentuan distribusi *prior* tersebut dilakukan berdasarkan *prior conjugate* yang terkait dengan pola *likelihood* data dimana λ dan γ dari distribusi Weibull merupakan keluarga distribusi eksponensial sehingga *prior*-nya adalah distribusi Gamma, sedangkan *prior* informatif yaitu prior berdasarkan informasi klasifikasi daerah perkotaan dan perdesaan. Distribusi *joint posterior* yang dihasilkan seperti pada persamaan (4.9) sehingga untuk mendapatkan distribusi *marginal posterior* dari masing-masing parameter dengan cara mengintegralkan keluar parameter-parameter yang bersangkutan sehingga dapat dijabarkan sebagai berikut,

$$p(\lambda_1 \mid \gamma_1, \pi_1) \propto \int_{\pi_1}\int_{\gamma_1} l\left(t \mid \gamma_1, \pi_1\right) p(\gamma_1) p(\pi_1)\ d\gamma_1 d\pi_1$$

$$p(\gamma_1 \mid \lambda_1, \pi_1) \propto \int_{\pi_1}\int_{\lambda_1} l\left(t \mid \lambda_1, \pi_1\right) p(\lambda_1) p(\pi_1)\ d\lambda_1 d\pi_1$$

$$p(\pi_1 \mid \lambda_1, \gamma_1) \propto \int_{\lambda_1}\int_{\gamma_1} l\left(t \mid \gamma_1, \lambda_1\right) p(\gamma_1)\, p(\lambda_1)\, d\gamma_1 d\lambda_1$$

$$\vdots$$

$$p(\pi_2 \mid \lambda_2, \gamma_2) \propto \int_{\lambda_2}\int_{\gamma_2} l\left(t \mid \gamma_2, \lambda_2\right) p(\gamma_2)\, p(\lambda_2)\, d\gamma_2 d\lambda_2$$

Dalam hal ini sampel yang digunakan sebanyak 11.960, dengan kondisi *burn in* mulai dari sampel 1001 dan *thin* 25 dengan iterasi sebanyak 300.000. Hasil estimasi parameter distribusi *mixture* ditunjukkan pada Tabel 4.10.

Tabel 4.10 Estimasi Parameter Distribusi *Mixture Weibull*

Parameter	Mean	Std. Dev	2,5%	Median	97,5%
Phi[1]	0,4012	0,01319	0,3752	0,401	0,4271
Phi[2]	0,5988	0,01319	0,5729	0,599	0,6248
pGamma[1]	2,009	0,07723	1,863	2,007	2,164
pGamma[2]	1,095	0,03486	1,029	1,094	1,165
pLambda[1]	0,0076	0,0015	0,0051	0,007	0,0108
pLambda[2]	0,0841	0,00737	0,0705	0,084	0,0991

Berdasarkan hasil estimasi pada Tabel 4.10 maka distribusi *mixture weibull* untuk data waktu *survival* lama sekolah penduduk dapat dinyatakan secara matematis sesuai persamaan (4.3) menjadi sebagai berikut:

$$p(t \mid \pi, \theta) = 0,4012(0,0076 \,\text{x}\, 2,009 t^{2,009-1}) \exp(-0,0076 t^{2,009}) + \\ 0,5988(0,0841 \,\text{x}\, 1,095 t^{1,095-1}) \exp(-0,00841 t^{1,095}),$$

Interpretasi dari model diatas yaitu bahwa distribusi lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat disusun oleh distribusi lama sekolah di daerah perkotaan sebesar 40,12 persen dan sebesar 59,88 persen untuk daerah perdesaan. Kontribusi yang lebih besar ditunjukkan di wilayah perdesaan dimana wilayah Provinsi Papua Barat didominasi oleh daerah perdesaan (Perka BPS Nomor 37 Tahun 2010).

Program untuk estimasi parameter distribusi *mixture weibull* beserta hasil estimasi secara lengkap dapat dilihat pada Lampiran 4 dan 8. Berdasarkan Lampiran 8 plot yang dihasilkan cenderung memiliki variansi kecil dan berjalan di sekitar rentang nilai tertentu, serta tidak ada yang memiliki jarak yang sangat jauh dan juga tidak berautokorelasi sehingga hasil estimasi yang diperoleh dapat dikatakan sudah cukup baik.

Fungsi *survival* dan fungsi *hazard* model *mixture weibull* dapat diperoleh melalui perhitungan sesuai persamaan (4.4) dan (4.5) dengan

$$\hat{S}(t) = 0,4012 \; \hat{S}_1(t) + 0,5988 \; \hat{S}_2(t),$$

dimana

$\hat{S}(t)$ = fungsi densitas untuk *mixture* lama sekolah di Provinsi Papua Barat

$\hat{S}_1(t)$ = fungsi densitas lama sekolah untuk daerah perkotaan

$\hat{S}_2(t)$ = fungsi densitas lama sekolah untuk daerah perdesaan.

Berdasarkan hasil estimasi parameter distribusi *weibull* angka lama sekolah penduduk pada Tabel 4.10 diperoleh hasil fungsi *survival* dan fungsi *hazard* sebagai berikut.

Tabel 4.11 Fungsi *Survival* dan Fungsi *Hazard* Model *Mixture*

t	$\hat{S}_1(t)$	$\hat{S}_2(t)$	$\hat{S}(t)$	$\hat{h}_1(t)$	$\hat{h}_2(t)$	$\hat{h}(t)$
1	0,9924	0,9193	0,9487	0,0153	0,0921	0,0613
2	0,9699	0,8356	0,8894	0,0307	0,0984	0,0712
3	0,9333	0,7557	0,8270	0,0463	0,1022	0,0798
4	0,8842	0,6813	0,7627	0,0618	0,1051	0,0877
5	0,8247	0,6126	0,6977	0,0775	0,1073	0,0953
6	0,7573	0,5498	0,6330	0,0931	0,1092	0,1027
7	0,6846	0,4925	0,5696	0,1088	0,1108	0,1100
8	0,6092	0,4405	0,5082	0,1245	0,1122	0,1171
9	0,5337	0,3935	0,4498	0,1402	0,1135	0,1242
10	0,4603	0,3511	0,3949	0,1559	0,1146	0,1312
11	0,3908	0,3129	0,3442	0,1716	0,1156	0,1381
12	0,3266	0,2786	0,2979	0,1874	0,1166	0,1450
13	0,2686	0,2478	0,2562	0,2031	0,1175	0,1519
14	0,2175	0,2203	0,2192	0,2189	0,1183	0,1587
15	0,1734	0,1956	0,1867	0,2347	0,1191	0,1655
16	0,1360	0,1736	0,1585	0,2505	0,1198	0,1722

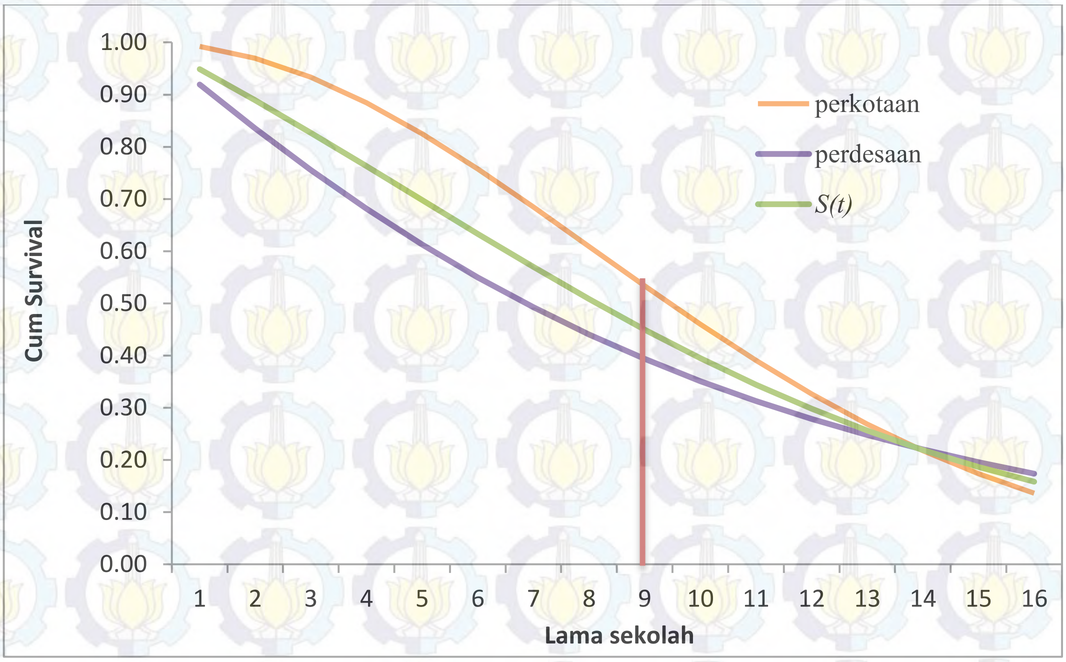
Dari hasil estimasi fungsi *survival* dan fungsi *hazard* model *mixture* pada Tabel 4.11 dapat diketahui bahwa nilai fungsi *survival* semakin menurun sementara nilai fungsi *hazard* semakin meningkat seiring lamanya waktu. Hal ini mengindikasikan bahwa semakin bertambah lamanya sekolah yang dijalani, maka kemungkinan seseorang untuk bertahan melanjutkan sekolah juga semakin rendah. Selain itu diperoleh pula informasi nilai fungsi *survival* dan fungsi *hazard* pada masing-masing komponen *mixture*.

Probabilitas penduduk dapat bertahan lanjut sekolah hingga 9 tahun yaitu sebesar 0,4498. Artinya persentase penduduk yang mampu bertahan untuk melanjutkan sekolah hingga tahun kesembilan yaitu sebesar 44,98 persen, belum mencapai setengah dari jumlah total penduduk pada rentang usia tersebut.

Selanjutnya fungsi *hazard* pada tahun kesembilan 0,1242 menunjukkan bahwa laju kegagalan seketika penduduk pada rentang usia 16-24 tahun setelah mampu bertahan untuk tetap bersekolah hingga mencapai pendidikan dasar 9 tahun adalah sebesar 12,42 persen. Nilai fungsi *hazard* semakin meningkat seiring dengan berjalannya waktu, dengan kata lain semakin lama penduduk melanjutkan sekolah maka kemungkinan untuk berhenti sekolah akan semakin tinggi.

Selain itu, Tabel 4.11 menunjukkan bahwa penduduk pada rentang usia 16-24 tahun di daerah perkotaan memiliki peluang untuk melanjutkan sekolah hingga mencapai pendidikan dasar 9 tahun yang lebih tinggi dibandingkan di daerah perdesaan, dimana di perkotaan sebesar 53,37 persen sementara di perdesaan hanya 39,35 persen. Sampai pada tahun ke-12 dimana seseorang telah berhasil mencapai pendidikan tingkat SMA, peluang penduduk pada rentang usia tersebut yang mampu bertahan melanjutkan sekolah hingga lulus SMA di perkotaan masih lebih tinggi dibandingkan dengan di perdesaan yakni 32,66 persen dan 27,86 persen masing-masing untuk di perkotaan dan perdesaan.

Gambar 4.11 merupakan grafik nilai fungsi *survival* untuk model *mixture* di daerah perkotaan dan perdesaan yang dihasilkan dari perhitungan pada Tabel 4.11.



Gambar 4.11. Plot Fungsi *Survival* Model *Mixture*

Gambar 4.11 menunjukkan plot fungsi *survival* model *mixture* serta kedua komponen *mixture*, yaitu daerah perkotaan dan perdesaan dimana terlihat bahwa probabilitas bertahan untuk melanjutkan sekolah pada daerah perkotaan lebih tinggi dibandingkan di daerah perdesaan. Tampak bahwa di daerah perkotaan penduduk pada rentang usia 16-24 tahun lebih besar peluangnya untuk melanjutkan sekolah hingga pendidikan 9 tahun dibandingkan penduduk dengan rentang umur yang sama di daerah perdesaan. Kecenderungan probabilitas yang lebih tinggi untuk tetap melanjutkan sekolah di daerah perkotaan dibandingkan di perdesaan masih konsisten hingga pada level pendidikan dasar 12 tahun (setingkat SMA). Namun pada tahun ke-14 dimana penduduk menempuh pendidikan kuliah tahun kedua, antara di perkotaan dan di perdesaan hampir sama peluangnya untuk tetap bertahan melanjutkan sekolah. Justru pada lama sekolah tahun berikutnya di daerah perdesaan menunjukkan peluang bertahan yang sedikit lebih tinggi untuk tetap melanjutkan pendidikan dibandingkan di daerah perkotaan. Fenomena seperti ini sangat mungkin terjadi karena ada sebagian kecil penduduk perdesaan yang memiliki minat belajar yang tinggi berusaha melanjutkan pendidikan ke sekolah-sekolah tinggi yang ada di perkotaan, tetapi bisa saja terjadi justru penduduk perkotaan lebih memilih tidak melanjutkan pendidikan ke jenjang yang lebih tinggi dengan alasan banyaknya lapangan pekerjaan di perkotaan yang lebih mudah didapat.

4.4.3. Pemodelan *Mixture Weibull Proportional Hazard*

Analisis *mixture survival* merupakan analisis *survival* dengan pendekatan distribusi *mixture* pada waktu *survival* (t). Seperti yang telah dijelaskan sebelumnya bahwa distribusi *mixture* yang digunakan untuk waktu *survival* lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat adalah distribusi *weibull* dengan dua komponen *mixture* yakni perkotaan dan perdesaan. Selanjutnya pemodelan untuk mengetahui faktor-faktor yang mempengaruhi penduduk Provinsi Papua Barat pada rentang usia tersebut dengan konsep distribusi *mixture* dapat disebut model *mixture weibull proportional hazard*. Bentuk umum model *mixture proportional hazard* seperti ditunjukkan pada persamaan (4.7) dan dilakukan estimasi parameter model *mixture survival* menggunakan metode *Bayesian* melalui teknik MCMC *gibbs sampling*.

Seperti halnya pada estimasi parameter *mixture*, sebelum melakukan estimasi terhadap parameter β (dalam syntak disimbolkan dengan beta), maka terlebih dahulu menentukan distribusi prior sebagai berikut,

$$\beta \sim Normal\left(e_{mk}, f_{mk}\right)$$

$$\gamma \sim Gamma\left(g_m, h_m\right)$$

dimana m = 1, 2 menentukan banyaknya komponen penyusun *mixture* dan k = 1, 2, …, p untuk banyaknya parameter β yang akan diestimasi. Penentuan distribusi *prior* untuk parameter β dilakukan berdasarkan *pseudo prior* yang diperoleh dari hasil estimasi parameter regresi Cox dengan metode *maximum likelihood* pada masing-masing komponen (yaitu daerah perkotaan dan perdesaan) melalui bantuan *software* SPSS 16 sehingga distribusi *prior* yang digunakan adalah distribusi normal. Begitu pula dengan penentuan distribusi *prior* untuk parameter γ, diperoleh berdasarkan hasil estimasi parameter *mixture* pada tahap sebelumnya sehingga didapatkan $\gamma \sim Gamma\left(g_m, h_m\right)$.

Distribusi *joint posterior* yang dihasilkan dapat dituliskan sebagai berikut,

$$p(\beta,\gamma,\pi \mid t,x,\delta) \propto l(t,x,\delta \mid \beta,\gamma,\pi)\times p(\beta)\times p(\gamma)\times p(\pi), \qquad (4.10)$$

dengan

$$l(t,x,\delta \mid \boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\pi}) = \prod_{i=1}^{n}\sum_{m=1}^{M}\pi_m\, f(t,x,\delta \mid \gamma_m,\beta_{mk},\pi_m),$$

dan simbol δ merupakan komponen indikator tersensor. Untuk mendapatkan distribusi *marginal posterior* dari masing-masing parameter dengan cara mengintegralkan keluar parameter-parameter yang bersangkutan sehingga dapat dijabarkan sebagai berikut,

$$p(\beta_{11} \mid \beta_{12},\beta_{13},\ldots,\beta_{16},\gamma_1,\pi_1) \propto \int_{\gamma_1}\int_{\pi_1}\int_{\beta_{12}}\cdots\int_{\beta_{16}} l(t \mid \gamma_1,\pi_1,\beta_{12},\beta_{13},\ldots,\beta_{16})\, p(\gamma_1)\, p(\pi_1)$$
$$p(\beta_{12})\cdots p(\beta_{16})\, d\gamma_1\, d\pi_1 d\beta_{12}\cdots d\beta_{16}$$

⋮

$$p(\beta_{26} \mid \beta_{21},\beta_{22},\ldots,\beta_{25},\gamma_2,\pi_2) \propto \int_{\gamma_2}\int_{\pi_2}\int_{\beta_{21}}\cdots\int_{\beta_{25}} l(t \mid \gamma_2,\pi_2,\beta_{21},\beta_{22},\ldots,\beta_{25})\, p(\gamma_2)\, p(\pi_2)$$
$$p(\beta_{21})\cdots p(\beta_{25})\, d\gamma_2\, d\pi_2 d\beta_{21}\cdots d\beta_{25}$$

Dalam hal ini sampel yang digunakan sebanyak 11.960, dengan kondisi *burn in* mulai dari sampel 1001 dan *thin* 25 dengan iterasi sebanyak 300.000 dengan bantuan *software* WinBUGS. Program untuk estimasi parameter distribusi *mixture weibull proportional hazard* beserta hasil estimasi secara lengkap dapat dilihat pada Lampiran 5 dan 9. Pada Lampiran 9, plot yang dihasilkan cenderung memiliki variansi kecil dan berjalan di sekitar rentang nilai tertentu, serta tidak ada yang memiliki jarak yang sangat jauh dan juga tidak berautokorelasi sehingga hasil estimasi yang diperoleh dapat dikatakan konvergen dan sudah cukup baik. Hasil estimasi parameter model *mixture weibull proportional hazard* ditunjukkan pada Tabel 4.12.

Tabel 4.12 Estimasi Parameter Model *Mixture Weibull Proportional Hazard*

Parameter	**Mean**	**exp (B)**	**Std.Dev**	**2,5%**	**Median**	**97,5%**
(1)	(2)	(3)	(4)	(5)	(6)	(7)
b1[1]	-0,0832	0,9202	0,0570	-0,1949	-0,0834	0,0272
b1[2]	-0,2816	0,7546	0,0535	-0,3860	-0,2812	-0,1764
b2[1]	-0,3921	0,6756	0,0624	-0,5158	-0,3918	-0,2701
b2[2]	-0,4617	0,6302	0,0568	-0,5722	-0,4618	-0,3496

Tabel 4.12 Estimasi Parameter Model *Mixture Weibull Proportional Hazard* (Lanjutan)

Parameter	**Mean**	**exp (B)**	**Std.Dev**	**2,5%**	**Median**	**97,5%**
(1)	(2)	(3)	(4)	(5)	(6)	(7)
b3[1]	-0,9758	0,3769	0,0693	-1,1110	-0,9759	-0,8386
b3[2]	-0,6626	0,5155	0,0635	-0,7864	-0,6625	-0,5378
b4_1[1]	0,1348	1,1443	0,0586	0,0186	0,1347	0,2481
b4_1[2]	0,1011	1,1064	0,0515	0,0002	0,1012	0,2016
b4_2[1]	-0,0501	0,9511	0,0637	-0,1734	-0,0508	0,0756
b4_2[2]	-0,0287	0,9717	0,0584	-0,1424	-0,0283	0,0865
b4_3[1]	0,0385	1,0393	0,0657	-0,0898	0,0385	0,1676
b4_3[2]	-0,0124	0,9877	0,0657	-0,1440	-0,0126	0,1194
b5[1]	-0,3251	0,7225	0,0177	-0,3594	-0,3252	-0,2901
b5[2]	-0,3783	0,6850	0,0171	-0,4125	-0,3781	-0,3453
b6[1]	-0,5411	0,5821	0,0500	-0,6397	-0,5405	-0,4431
b6[2]	-0,6546	0,5196	0,0610	-0,7740	-0,6545	-0,5331

Hasil pemodelan *mixture proportional hazard* pada waktu *survival* lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat pada Tabel 4.12 menunjukkan bahwa tidak semua variabel yang digunakan dalam penelitian ini signifikan memberikan pengaruh terhadap lama sekolah penduduk. Berdasarkan hasil pemodelan di atas maka variabel yang dinyatakan memberi pengaruh yang signifikan terhadap angka lama sekolah di perkotaan dengan selang kepercayaan 95% antara lain: status bekerja (X_2), status perkawinan (X_3), pendidikan KRT tidak tamat SD (X_{41}), jumlah ART (X_5) dan rata-rata pengeluaran rumah tangga (X_6). Sementara di perdesaan, variabel yang dinyatakan berpengaruh terhadap angka lama sekolah dengan selang kepercayaan 95% adalah jenis kelamin (X_1), status bekerja (X_2), status perkawinan (X_3), pendidikan KRT tidak tamat SD (X_{41}), jumlah ART (X_5) dan rata-rata pengeluaran rumah tangga (X_6).

Model hubungan faktor-faktor yang berpengaruh terhadap lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat berdasarkan persamaan (4.7) sesuai dengan hasil estimasi parameter *mixture weibull proportional hazard* untuk daerah perkotaan secara matematis dapat dituliskan sebagai berikut:

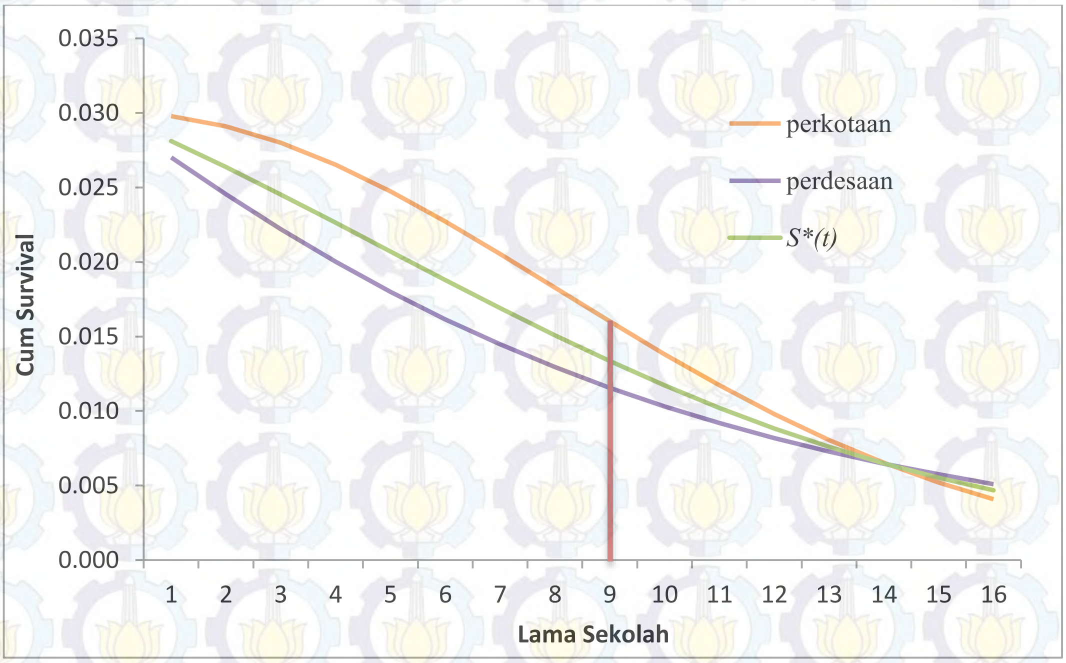
$$h_{i1}(t) = 0,4012 \text{ x} (0,0076 \text{ x} 2,009t^{2,009-1}) \exp(-0,3921X_{2i} - 0,9758X_{3i} + 0,1348X_{41i} - 0,3251X_{5i} - 0,5411X_{6i}),$$

dan model untuk daerah perdesaan adalah:

$$h_{i2}(t) = 0,5988 \text{ x} (0,0841 \text{ x} 1,095t^{1,095-1}) \exp(-0,2816X_{1i} - 0,4617X_{2i} - 0,6626X_{3i} + 0,1011X_{41i} - 0,3783X_{5i} - 0,6546X_{6i}).$$

Dari hasil pemodelan diatas dapat disimpulkan bahwa distribusi lama sekolah penduduk usia 16 – 24 tahun di Papua Barat disusun oleh distribusi lama sekolah di daerah perkotaan sebesar 40,12 persen dan sebesar 59,88 persen merupakan kontribusi dari daerah perdesaan.

Selanjutnya, dengan diberikannya variabel-variabel prediktor yang berpengaruh secara signifikan terhadap angka lama sekolah di masing-masing daerah dan diperoleh model *mixture weibull proportional hazard*, fungsi survival yang terbentuk berdasarkan persamaan (4.8) dapat disajikan seperti pada Gambar 4.12.



Gambar 4.12. Plot Fungsi *Survival* Model *Mixture Weibull Proportional Hazard*

Gambar 4.12 menunjukkan plot fungsi *survival* model *mixture* weibull *proportional hazard* dimana melibatkan variabel-variabel yang berpengaruh secara signifikan untuk masing-masing komponen *mixture*, yaitu daerah perkotaan dan perdesaan. Terlihat bahwa probabilitas bertahan untuk melanjutkan sekolah pada daerah perkotaan justru lebih tinggi dibandingkan di daerah perdesaan. Tampak bahwa di daerah perkotaan penduduk pada rentang usia 16-24 tahun lebih tinggi peluangnya untuk melanjutkan sekolah hingga pendidikan 9 tahun dibandingkan penduduk dengan rentang umur yang sama di daerah perdesaan. Kecenderungan probabilitas yang lebih tinggi untuk tetap melanjutkan sekolah di daerah perkotaan dibandingkan di perdesaan masih konsisten hingga pada level SMA. Namun uniknya pada level pendidikan dua tahun kuliah/perguruan tinggi (lama sekolah 14 tahun) pada kedua daerah memiliki peluang bertahan untuk melanjutkan sekolah yang sama besarnya walaupun sangat kecil nilainya yakni sekitar 0,0065 dan pada saat mencapai pendidikan 15 tahun peluang ketahanan sekolah untuk di perdesaan sedikit lebih tinggi dibandingkan di perkotaan.

Hasil tabulasi analisis secara deskriptif terkait karakteristik penduduk dengan angka lama sekolah 14 tahun ke atas seperti ditunjukkan pada Lampiran 9. Perbandingan penduduk pada rentang usia 16 – 24 tahun dengan angka lama sekolah 14 tahun ke atas yang tidak bekerja terhadap yang bekerja di perkotaan lebih rendah daripada di perdesaan. Artinya di perkotaan, penduduk dengan angka lama sekolah 14 tahun ke atas cukup banyak yang berhenti sekolah karena mendapatkan pekerjaan. Sementara itu, di perdesaan terdapat 26,87 persen penduduk yang angka lama sekolahnya 14 tahun ke atas dengan status bekerja. Dari sisi pendidikan kepala rumah tangga (KRT), penduduk dengan angka lama sekolah 14 tahun ke atas di perkotaan yang KRT-nya tidak tamat SD memiliki persentase yang lebih rendah dibandingkan di perdesaan. Hal ini menunjukkan bahwa belum tentu angka lama sekolah seseorang dipengaruhi secara langsung dari tingkat pendidikan KRT. Namun di satu sisi, secara ekonomi penduduk pada rentang usia tersebut yang memiliki angka lama sekolah 14 tahun ke atas di perkotaan masih lebih tinggi besar rata-rata pengeluaran perkapita per bulannya di bandingkan di perdesaan. Ini artinya tidak selalu rumah tangga dengan pendapatan yang tinggi akan bersekolah lebih tinggi dibandingkan dengan rumah tangga dengan pendapatan rendah. Rumah tangga dengan pengeluaran perkapita yang tinggi tidak selalu menginvestasikannya untuk pendidikan. Hal inilah yang menyebabkan pola kurva fungsi ketahanan bersekolah di perkotaan tampak sedikit lebih rendah dibandingkan di perdesaan pada angka lama sekolah 14 tahun ke atas.

Adanya fenomena penduduk di perkotaan memilih untuk bekerja dibandingkan melanjutkan pendidikan ke jenjang yang lebih tinggi karena tingginya peluang bekerja di daerah perkotaan. Di daerah perkotaan dewasa ini mulai berkembang usaha-usaha di sektor perdagangan, pelayanan dan jasa serta sektor yang lain. Sementara itu, penduduk di daerah perdesaan sedikit demi sedikit mulai beradaptasi dengan pengaruh masyarakat perkotaan serta hadirnya pendatang di daerahnya sehingga ada sebagian kecil diantara mereka yang tetap bertahan untuk melanjutkan pendidikan ke jenjang yang lebih tinggi.

Hal ini menunjukkan bahwa variabel-variabel yang berpengaruh secara signifikan terhadap angka lama sekolah pada masing-masing wilayah menghasilkan estimasi peluang bahwa penduduk usia 16 – 24 tahun di daerah perkotaan lebih lama untuk bertahan melanjutkan sekolah daripada di daerah perdesaan.

Interpretasi model *mixture weibull proportional hazard* untuk daerah perkotaan dapat dijelaskan sebagai berikut:

a. Variabel status bekerja (X_2) menunjukkan bahwa nilai $\hat{\beta}_2$= -0,3921 sehingga $\exp(\hat{\beta}_2)$= 0,6756. Artinya resiko penduduk yang bekerja lebih lama sekolahnya adalah sebesar 0,6756 kalinya penduduk yang tidak bekerja. Dengan kata lain resiko bahwa penduduk di daerah perkotaan yang tidak bekerja dapat melanjutkan pendidikan lebih lama yakni 1,4802 kalinya (1/0,6756) penduduk yang statusnya bekerja. Hal ini didukung oleh teori yang diungkapkan oleh Ersado (2005) dalam penelitiannya di Nepal, Peru dan Zimbabwe yang membahas tentang imbas pekerja anak terhadap kemungkinan seorang anak untuk tetap bersekolah.
b. Variabel status perkawinan (X_3) menunjukkan bahwa nilai $\hat{\beta}_3$= -0,9758 sehingga diperoleh nilai $\exp(\hat{\beta}_3)$ sebesar 0,3769, yang artinya resiko penduduk

usia 16 – 24 tahun di perkotaan yang tidak kawin adalah sebesar 0,3769 kalinya penduduk pada rentang usia tersebut yang berstatus kawin. Atau dapat diartikan bahwa resiko penduduk di daerah perkotaan yang berstatus kawin dapat melanjutkan pendidikan lebih lama yaitu 2,6532 kalinya (1/0,3769) penduduk yang statusnya tidak kawin. Pada kenyataannya di Provinsi Papua Barat banyak juga ditemukan anak usia sekolah yang statusnya sudah kawin (secara adat/agama) walaupun masih setingkat SMA. Konsep perkawinan yang ada di Susenas yaitu ketika seseorang mempunyai istri (bagi laki-laki) atau suami (bagi perempuan) pada saat pencacahan, baik tinggal bersama maupun terpisah dimana mencakup perkawinan yang sah secara hukum (adat, agama, negara dan sebagainya) dan juga mereka yang hidup bersama dan oleh masyarakat sekelilingnya dianggap sebagai suami-istri. Namun hal ini bukan berarti mengindikasikan bahwa lebih baik perkawinan di usia dini memberikan dampak yang positif terhadap pencapaian angka lama sekolah, tetapi lebih ditekankan pada kualitas lulusan siswa dalam persaingan di dunia kerja.

c. Variabel pendidikan KRT tidak tamat SD (X_{41}) menunjukkan bahwa nilai $\hat{\beta}_{41}$ = 0,1348 sehingga didapatkan nilai $\exp(\hat{\beta}_{41})$ sebesar 1,1443, yang artinya penduduk usia 16 – 24 tahun di perkotaan dimana kepala rumah tangganya tidak tamat SD memiliki resiko bersekolah lebih lama 1,1443 kalinya penduduk pada rentang usia tersebut yang kepala rumah tangganya tamat minimal SD. Hal ini tidak sejalan dengan penelitian yang dilakukan oleh Brunello dan Checchi pada tahun 2005 tentang pengaruh tingkat pendidikan orang tua terhadap tingkat pencapaian pendidikan penduduk. Hal ini bisa disebabkan karena adanya dorongan yang kuat dari diri seseorang untuk mencapai pendidikan yang lebih baik dibandingkan yang telah diperoleh orang tuanya guna memperbaiki kehidupan yang lebih mapan. Khususnya di daerah perkotaan, semakin banyak pendatang yang bermukim di Papua Barat dengan menciptakan hasil karya yang lebih baik maka mendorong masyarakat untuk menempuh pendidikan yang lebih tinggi.

d. Variabel jumlah ART (X_5) menunjukkan nilai estimasi $\hat{\beta}_5$ yang negatif yaitu sebesar -0,3251 sehingga hubungan dengan angka lama sekolah negatif, maka didapatkan nilai $\exp(\hat{\beta}_5)$ sebesar 0,7225, yang artinya setiap penambahan 1 orang anggota rumah tangga pada rumah tangga penduduk usia 16 – 24 tahun di perkotaan lebih cenderung menurunkan resiko lama sekolah yang akan dicapai hingga 1,3841 kalinya (1/0,7225) kalinya. Hasil ini sesuai dengan penelitian yang diuraikan oleh Lloyd (1996) yang dilakukan di wilayah *Sub Saharan Africa* dimana keluarga yang memiliki jumlah anggota rumah tangga lebih banyak maka pencapaian tingkat pendidikan yang berhasil ditempuh tidak lebih tinggi jika dibandingkan dengan keluarga yang memiliki jumlah anggota rumah tangga yang sedikit.
e. Variabel rata-rata pengeluaran rumah tangga sebulan (X_6) menunjukkan nilai estimasi $\hat{\beta}_6$ yang negatif yaitu sebesar -0,5411 sehingga hubungan dengan angka lama sekolah negatif, maka didapatkan nilai $\exp(\hat{\beta}_6)$ sebesar 0,5821. Ini berarti bahwa peningkatan 1 juta rupiah pengeluaran perkapita per bulan justru menurunkan resiko penduduk usia 16 – 24 tahun di perkotaan untuk melanjutkan pendidikan yang lebih tinggi. Dengan kata lain, rumah tangga perkotaan yang memiliki pendapatan lebih tinggi belum tentu mengalokasikan pengeluaran rumah tangga untuk investasi pendidikan yang lebih besar dibandingkan dengan rumah tangga pendapatan yang lebih rendah. Berbeda dengan penelitian-penelitian sebelumnya seperti Blanden dan Gregg (2004) yang menunjukkan adanya korelasi positif antara tingkat pendapatan dengan pencapaian pendidikan anak, penduduk wilayah perkotaan di Provinsi Papua Barat justru memiliki kecenderungan sebaliknya. Hal ini diperkirakan karena prioritas konsumsi non-makanan rumah tangga (seperti alokasi pengeluaran pendidikan dan kesehatan) lebih rendah dibandingkan dengan konsumsi makanan rumah tangga. Hal ini didukung dengan hasil pendataan Susenas 2014 di Papua Barat yang dipublish dalam publikasi Inkesra 2014 bahwa di beberapa kabupaten cenderung memiliki persentase pengeluaran makanan yang lebih tinggi dibandingkan non-makanan.

Selanjutnya, interpretasi model *weibull mixture weibull proportional hazard* untuk daerah perdesaan dapat dijelaskan sebagai berikut:

a. Variabel jenis kelamin (X_1) menunjukkan bahwa nilai $\hat{\beta}_1$ = -0,2816 sehingga $\exp(\hat{\beta}_1)$ = 0,7546. Artinya resiko penduduk laki-laki lebih lama sekolahnya adalah 0,7546 kalinya penduduk perempuan. Atau dengan kata lain resiko bahwa penduduk perempuan dapat melanjutkan pendidikan lebih lama yaitu 1,3252 kali (1/0,7546) penduduk laki-laki pada rentang usia tersebut. Jika diperbandingkan dalam bentuk jumlah, maka dapat diinterpretasikan bahwa ketika terdapat 100 orang penduduk laki-laki berusia antara 16 – 24 tahun di daerah perdesaan yang mampu menyelesaikan pendidikan dasar 9 tahun, maka ada sekitar 132 orang penduduk berjenis kelamin perempuan yang mampu menyelesaikan pendidikan dasar 9 tahun.

b. Variabel status bekerja (X_2) menunjukkan bahwa nilai $\hat{\beta}_2$= -0,4617 sehingga $\exp(\hat{\beta}_2)$= 0,6302. Artinya resiko penduduk yang bekerja lebih lama sekolahnya adalah sebesar 0,6302 kalinya penduduk yang tidak bekerja. Dengan kata lain resiko bahwa penduduk di daerah perdesaan yang tidak bekerja dapat melanjutkan pendidikan lebih lama yakni 1,5868 kalinya (1/0,6302) penduduk yang statusnya bekerja. Sama halnya dengan di perkotaan, hasil ini didukung oleh teori yang diungkapkan oleh Ersado (2005) dalam penelitiannya di Nepal, Peru dan Zimbabwe yang membahas tentang imbas pekerja anak terhadap kemungkinan seorang anak untuk tetap bersekolah.

c. Variabel status perkawinan (X_3) menunjukkan bahwa nilai $\hat{\beta}_3$= -0,6626 sehingga diperoleh nilai $\exp(\hat{\beta}_3)$ sebesar 0,5155, yang artinya resiko penduduk usia 16 – 24 tahun di perdesaan yang tidak kawin adalah sebesar 0,5155 kalinya penduduk pada rentang usia tersebut yang berstatus kawin. Atau dapat diartikan bahwa resiko penduduk di daerah perdesaan yang berstatus kawin dapat melanjutkan pendidikan lebih lama yaitu 1,9399 kalinya (1/0,5155) penduduk yang statusnya tidak kawin. Khususnya di daerah pedalaman di Provinsi Papua Barat masih ditemukan anak usia sekolah yang masih bersekolah sembari mengasuh anak. Meskipun begitu, kualitas lulusan yang dihasilkan di perdesaan bisa dikatakan cukup rendah untuk dipersaingkan dengan di perkotaan.

d. Variabel pendidikan KRT tidak tamat SD (X_{41}) menunjukkan bahwa nilai $\hat{\beta}_{41}$ = 0,1011 sehingga didapatkan nilai $\exp(\hat{\beta}_{41})$ sebesar 1,1064, yang artinya penduduk usia 16 – 24 tahun di daerah perdesaan dimana kepala rumah tangganya tidak tamat SD memiliki resiko bersekolah lebih lama 1,1064 kalinya penduduk pada rentang usia tersebut yang kepala rumah tangganya tamat minimal SD. Meski tidak sejalan dengan penelitian yang dilakukan oleh Schultz (2002) tentang keterkaitan antara tingkat pendidikan orang tua keinginan untuk menyekolahkan anak-anak mereka, pengaruh yang muncul akibat adanya migrasi penduduk luar daerah yang datang ke wilayah perdesaan diperkirakan sedikit demi sedikit mulai mengubah pola pikir penduduk untuk lebih maju.

e. Variabel jumlah ART (X_5) menunjukkan nilai estimasi $\hat{\beta}_5$ yang negatif yaitu sebesar -0,3783 sehingga hubungan variabel ini dengan angka lama sekolah negatif, maka didapatkan nilai $\exp(\hat{\beta}_5)$ sebesar 0,685, yang artinya setiap penambahan 1 orang anggota rumah tangga pada rumah tangga penduduk usia 16 – 24 tahun di daerah perdesaan cenderung menurunkan resiko lama sekolah yang akan dicapai hingga 1,4598 kalinya (1/0,685) kalinya. Sama halnya dengan di perkotaan, bahwa hasil ini seiring dengan penelitian yang dilakukan oleh Lloyd (1996) di wilayah *Sub Saharan Africa* dimana keluarga yang memiliki jumlah anggota rumah tangga lebih banyak maka pencapaian tingkat pendidikan yang berhasil ditempuh tidak lebih tinggi jika dibandingkan dengan keluarga yang memiliki jumlah anggota rumah tangga yang sedikit.

f. Variabel rata-rata pengeluaran rumah tangga sebulan (X_6) menunjukkan nilai estimasi $\hat{\beta}_6$ yang negatif yaitu sebesar -0,6546 sehingga hubungan dengan angka lama sekolah negatif, maka didapatkan nilai $\exp(\hat{\beta}_6)$ sebesar 0,5196. Ini artinya setiap peningkatan 1 juta rupiah pengeluaran perkapita per bulan justru menurunkan resiko penduduk usia 16 – 24 tahun di perdesaan untuk melanjutkan pendidikan yang lebih tinggi. Atau dengan kata lain, rumah tangga di perdesaan yang memiliki pendapatan lebih rendah bisa jadi lebih banyak mengalokasikan pengeluaran rumah tangga untuk investasi pendidikan sehingga tingkat pendidikan yang mampu dicapai anaknya lebih baik dibandingkan dengan rumah tangga yang pendapatannya lebih tinggi namun investasi pendidikan yang dikeluarkan lebih sedikit.

Dari hasil di atas dapat disimpulkan bahwa kontribusi di daerah perkotaan lebih kecil daripada kontribusi dari daerah perdesaan. Akibatnya, peluang bertahan untuk bersekolah di Papua Barat akan sangat besar pengaruhnya yang disebabkan karena besarnya kontribusi dari daerah perdesaan. Selain itu, tampak adanya perbedaan faktor yang berpengaruh terhadap angka lama sekolah penduduk usia 16 – 24 tahun antara di daerah perkotaan dan di daerah perdesaan. Berdasarkan variabel-variabel yang berpengaruh secara signifikan pada masing-masing daerah dapat diambil kesimpulan bahwa peluang untuk melanjutkan sekolah pada penduduk usia 16 – 24 tahun di perdesaan lebih rendah daripada di daerah perkotaan. Misalnya dibandingkan di perkotaan, resiko penduduk pada rentang usia tersebut di daerah perdesaan yang tidak bekerja untuk tetap bersekolah masih lebih tinggi yaitu bisa mencapai 1,5 kalinya dari yang bekerja. Ini menunjukkan bahwa efek seseorang bekerja di daerah perdesaan memiliki resiko yang lebih besar terhadap peningkatan angka lama sekolah dibandingkan di perkotaan.

Selain variabel-variabel yang berpengaruh dalam model diatas,perbedaan pencapaian sekolah antara kedua daerah bisa juga disebabkan karena minimnya fasilitas pendidikan di wilayah perdesaan, terbatasnya tenaga pendidik, serta keterbatasan ditinjau dari segala aspek pembangunan fasilitas umum lainnya. Dengan kondisi keterbatasan fasilitas yang ada, masih tampak adanya semangat yang cukup tinggi penduduk usia tersebut untuk tetap bersekolah yang ditunjukkan dari fungsi ketahanan melanjutkan pendidikan yang lebih tinggi. Adanya tekad untuk membangun daerah untuk menjadi lebih baik selayaknya perlu mendapatkan dukungan serta perhatian yang serius dari pemerintah khususnya pemerintah daerah.

BAB 5
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian ini, dengan menggunakan model *mixture weibull proportional hazard* dengan pendekatan *Bayesian* diperoleh kesimpulan sebagai berikut:

1. Kontribusi yang dihasilkan dari distribusi *mixture weibull* dari angka lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat yaitu sebesar 40,12 persen di daerah perkotaan, dan 59,88 persen di daerah perdesaan.
2. Dari model *mixture proportional hazard* dihasilkan bahwa peluang penduduk usia 16 – 24 tahun di daerah perkotaan untuk dapat melanjutkan sekolah atau pendidikan dasar 9 tahun lebih tinggi daripada di daerah perdesaan.
3. Variabel yang memberikan pengaruh secara signifikan terhadap angka lama sekolah di daerah perkotaan berdasarkan model *mixture weibull proportional hazard* berbeda dengan di daerah perdesaan dimana di daerah perkotaan variabel jenis kelamin tidak berpengaruh secara signifikan namun berpengaruh signifikan di daerah perdesaan.

5.2. Saran

Dengan memperhatikan beberapa hal yang telah disimpulkan di atas, maka yang dapat disarankan untuk penelitian selanjutnya adalah:

1. Sebagai bahan pertimbangan pada Pemerintah Daerah Provinsi Papua Barat, bahwa terdapat perbedaan yang jelas antara ketahanan penduduk untuk melanjutkan sekolah ke jenjang yang lebih tinggi dimana perkotaan cenderung lebih tinggi. Namun kontribusi terbesar berasal dari perdesaan. Hal ini tentunya memberikan pengaruh terhadap angka lama sekolah di Provinsi Papua Barat. Oleh karenanya, perlu adanya kebijakan terkait pengaruh pekerja anak terhadap angka lama sekolah baik di daerah perkotaan maupun perdesaan.

2. Dari sisi perencanaan keluarga, menjadi suatu agenda penting tentang adanya pencanangan program keluarga berencana sebagai salah satu bentuk upaya agar dapat menghasilkan generasi muda yang lebih cemerlang.
3. Selain itu, perlu juga adanya sosialisasi kepada masyarakat tentang pentingnya manajemen pengeluaran konsumsi rumah tangga, seperti investasi pendidikan anak untuk masa depan yang lebih baik.
4. Untuk penelitian selanjutnya, perlu dipertimbangkan adanya faktor dari karakteristik wilayah seperti ketersediaan fasilitas sekolah, kuantitas dan kualitas tenaga pendidik, dan lain sebagainya.

DAFTAR PUSTAKA

Aksioma, D. F. (2011), *Model Spasial Survival dengan Pendekatan Bayesian (Studi Kasus pada Kejadian HIV/AIDS di Provinsi Jawa Timur),* Surabaya: Institut Teknologi Sepuluh Nopember.

Badan Pusat Statistik (2010), *Peraturan Kepala BPS Nomor 37 Tahun 2010 tentang Klasifikasi Perkotaan dan Perdesaan*, BPS: Jakarta.

Badan Pusat Statistik (2015), *Indeks Pembangunan Manusia Metode Baru*, BPS: Jakarta.

BKKBN (2011), Kajian Profil Penduduk Remaja (10 – 24 Thn): Ada apa dengan Remaja?, *Policy Brief*, Seri I No. 6.

Blanden, J. dan Gregg, P. (2004), Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain, *CMPO Working Paper*, 04/101.

Box, G.E.P. dan Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis,* Addison-Wesley, London.

Brunello, G. dan Checchi, D. (2005), "School quality and family background in Italy", *Economics of Education Review,* 24, 563 – 577.

Carlin, B.P. dan Chib, S. (1995), "*Bayesian* Model Choice via Markov Chain Monte Carlo Methods", *Journal of The Royal Statistical Society,* 57(3), 473 - 484.

Casella, G. dan George, E.I. (1992), "Explaining Gibbs Sampler", *The America Statistical Association,* 46(3), 167-174.

Collet, D. (1994), *Modelling Survival Data in Medical Research,* Chapman and Hall, London.

Collet, D. (2003). *Modelling Survival Data in Medical Research 2nd ed.* London: Chapman and Hall.

Cox, D.R. (1972), “Regression Models and Life Tables”, *Journal of The Royal Statistical Society,* 34, 187-220.

Cox, D., dan Oakes, D. (1984). *Analysis of Survival Data.* London: Chapman and Hall.

Darmofal, D. (2008), *Bayesian Spatial Survival Models for Political Event Processes.* Columbia: Departement of Political Science, University of South Carolina 350 Gambrel Hall.

Delprato, M., Akyeampong, K., Sabates, R., dan Hernandes-Fernandes, J. (2015), On the impact of early marriage on schooling outcomes in Sub-Saharan Africa and South West Asia, *International Journal of Educational Development*, 44, 42 – 55.

Ersado, L. (2005), Child Labor and Schooling Decisions in Urban and Rural Areas: Comparative Evidence from Nepal, Peru, and Zimbabwe, *World Development*, Vol. 33(3), 455 – 480.

Gamerman, D. (1997), *Markov Chain Monte Carlo,* Chapman & Hall, London.

Gelman, A., Carlin, J.B., Stern H.S., Rubin D.B. (1995), *Bayesian Data Analysis, 2nd Edition,* Chapman and Hall, London.

Hariyanto, S. (2009), *Model Mixture Survival pada Kasus Lama Mencari Kerja di Pulau Jawa Tahun 2007*, Tesis, Surabaya: Institut Teknologi Sepuluh Nopember.

Hiliry, M.D. (1995), *Wanita Usia 7-12 Tahun yang Tidak bersekolah dan Faktor Penyebabnya di Daerah Istimewa Aceh dalam Laporan Akhir: Anak Perempuan Usia 7-12 Tahun yang Tidak Sekolah dan Putus Sekolah di Sumatera*, United Nations Fund for Population Activities (UNFPA) dengan Pusat Studi Kependudukan Universitas Andalas, Padang.

Hosmer Jr., D.W. dan Lemenshow, S. (1999), *Applied Survival Analysis: Regression Modelling of Time to Event Data.* John Wiley and Sons. Inc., New York.

Hurn, M., Justel, A., dan Robert, C.P., (2000), *Estimating Mixture of Regressions*, CREST, Insee, Paris.

Ikana, M. (2005). *Pengaruh Urutan Kelahiran Terhadap Kelangsungan Pendidikan Anak Perempuan Usia 7-15 Tahun di Indonesia (Analisis Data Susenas 2002-KOR),* Tesis. Depok: Universitas Indonesia.

Iriawan, N. (2000), *Computationally Intensive Approaches to Inference in Neo-Normal Linier Models,* Thesis Ph.D., CUT-Australia.

Iriawan, N. (2001), *Studi tentang Bayesian Mixture Normal dengan Menggunakan Metode MCMC,* Laporan penelitian: Lemlit ITS, Surabaya.

Iriawan, N. (2002), *Studi Tentang Model Mixtures Regresi Linear, Pendekatan Markov Chain Monte Carlo (MCMC)*, *Natural Journal*, 249-256.

Iriawan, N. dan Astuti, S.P. (2006), *Mengolah Data Statistik dengan Mudah Menggunakan Minitab 14*, Andi Offset, Yogyakarta.

Kleinbaum, D.G. dan Klein, M. (2005), *Survival Analysis: A Self Learning, 2nd Edition,* Springer, New York.

Law, A.M., dan Kelton D.W. (2000). *Simulation Modeling Analysis (3rd Edition)*. New York: MacGraw-Hill.

Lawless, J.F. (2003), *Statistical Models and Methods for Lifetime Data 2nd Edition,* John Wiley and Sons. Inc., New York.

Lee, E. (1992). *Statistical Models and Methods for Lifetime Data.* New York: John Wiley and Sons. Inc.

Lloyd, C. B. dan Blanc, A. K. (1996), "Children's Schooling in Sub-Saharan Africa: The Role of Fathers, Mothers and Others". *Population and Development Review*, 22(2), 265 – 98.

Marin, J.M., Mengersen, K., dan Robert, C.P. (2001), "*Bayesian* Modelling and Inference on *Mixture* of Distribution", *Handbook of Statistics*, 25, 50.

McCullagh, P., dan Nelder, J.A. (1989), *Generalized Linear Models (Second Edition),* Chapman and Hall, New York.

McLachlan, G., dan Basford, K.E. (1988), *Mixture Models Inference and Applications to Clustering,* Marcel Dekker, New York.

McLachlan, G., dan Peel, D. (2000), *Finite Mixture Models*, John Wiley and Sons. Inc., New York.

McMahon W., (1999), *Education and Development: Measuring the Social Benefits*, Oxford University Press, Oxford.

Miller, R. (1998). *Survival Analysis.* New York: John Willey and Sons Inc.

Muthen, B. dan Masyn, K. (2005), "Discrete-Time *Mixture Survival* Analysis", *Journal of Educational and Behavioral Statistics*, 30(1), 27-58.

Ntzoufras, I. (2009). *Bayesian Modelling Using WinBUGS.* New York: John Willey and Sons, Inc.

Qian, I. (1994). *A Bayesian Weibull Survival Model*. Disertation. Institute of Statistics and Decision Science in Graduate School of Duke University.

Qudsi, J. (2015). *Model Mixture Survival Spasial pada angka Lama Sekolah Anak Umur 16-18 Tahun Di Jawa Timur Tahun 2012*, Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.

Santoso, B. (2009). *Pendekatan Spline Multivariable Dan Mars Untuk Pemodelan Lama Sekolah Pada Penduduk Usia Sekolah Di Provinsi Papua,* Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.

Schultz, T. Paul, (2002), "Why Governments Should Invest More to Educate Girls," *World Development*, 30: 207-225

Setyawan, N.A.D. (2011). *Pendekatan Regresi Nonparametrik Birespon Spline untuk Pemodelan Determinan Tingkat Pendidikan di Pulau Papua*, Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.

Solikhah, A. (2009). *Analisis Rata-Rata Lama Sekolah di Pulau Kalimantan Menggunakan Model Spasial Conditional Autoregression (CAR)*, Tesis. Surabaya: Institut Teknologi Sepuluh Nopember.

Suendra, N. (1999). *Pemetaan Peserta Didik Wajib Belajar 9 Tahun yang Putus Sekolah dan Tidak Melanjutkan Ke SLTP Pada Desa-Desa Tertinggal di Provinsi Bali, Tugas Akhir*. Bali: Jurusan Ilmu Pendidikan dan Keguruan Universitas Singaraja.

Stephen, M. (1997), *Bayesian Metods for Mixture of Normal Distribution*. Thesis, University of Oxford, UK.

Sulistiyawati, D. (2009). *Model Mixture Survival Pada Kasus Lama Sekolah di Kabupaten Boalemo Provinsi Gorontalo, Tesis.* Surabaya: Institut Teknologi Sepuluh Nopember.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2005). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.* New York: Springer.

Zang. (2008). *Survival Analysis.* California: Wadsworth.

halaman ini sengaja dikosongkan

LAMPIRAN 1 Statistik Deskriptif

Untuk Daerah Perkotaan

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Lama Sekolah	559	.00	16.00	11.0823	2.71271
jumlah ART	559	2.00	18.00	5.8945	2.49256
Rata2 Pengeluaran Perkapita 1 bln	559	.26	6.00	1.0637	.71331
Valid N (listwise)	559				

Untuk Daerah Perdesaan

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Lama Sekolah	835	.00	16.00	9.4311	3.41895
jumlah ART	835	2.00	20.00	5.9186	2.47541
Rata2 Pengeluaran Perkapita 1 bln	835	.14	8.75	.5572	.46383
Valid N (listwise)	835				

LAMPIRAN 2. Uji Distribusi Data Lama Sekolah

***Goodness of Fit* – Provinsi Papua Barat**

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gamma	0.18568	5	106.06	5	610.56	6
2	Gamma	0.16888	4	99.292	4	624.35	7
3	Log-Logistic	0.28638	9	222.64	9	1875.3	9
4	Log-Logistic	0.13525	2	97.523	3	517.9	5
5	Lognormal	0.19803	7	120.47	7	484.83	3
6	Lognormal	0.19801	6	120.47	6	484.83	4
7	Normal	0.14901	3	25.536	1	143.96	1
8	Weibull	0.24214	8	219.56	8	1689.7	8
9	Weibull	0.12605	1	74.28	2	264.67	2

***Goodness of Fit* – Daerah Perkotaan**

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gamma	0.13944	3	21.998	4	278.23	7
2	Gamma	0.13804	2	22.341	5	234.88	5
3	Log-Logistic	0.26771	9	61.422	9	382.31	8
4	Log-Logistic	0.13538	1	18.36	3	240.05	6
5	Lognormal	0.16522	6	29.102	6	97.75	3
6	Lognormal	0.16523	7	29.104	7	97.745	2
7	Normal	0.14217	4	9.4611	1	225.73	4
8	Weibull	0.22312	8	59.735	8	428.62	9
9	Weibull	0.16221	5	16.566	2	95.5	1

LAMPIRAN 2. Uji Distribusi Data Lama Sekolah (Lanjutan)

***Goodness of Fit* – Daerah Perkotaan**

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gamma	0.19804	7	76.535	5	206.67	3
2	Gamma	0.16747	4	71.358	3	253.73	4
3	Log-Logistic	0.26236	9	147.13	9	1169.7	9
4	Log-Logistic	0.13657	2	75.025	4	421.07	7
5	Lognormal	0.1928	6	83.744	7	337.45	5
6	Lognormal	0.19279	5	83.742	6	337.45	6
7	Normal	0.15042	3	14.012	1	190.52	2
8	Weibull	0.25653	8	146.39	8	876.04	8
9	Weibull	0.11199	1	57.506	2	139.36	1

LAMPIRAN 3. Syntax Regresi Cox *Proportional Hazard* dengan SPSS

Untuk Daerah Perkotaan

```
USE ALL.
 COMPUTE filter_$=(x7=1).
 VARIABLE LABEL filter_$ 'x7=1 (FILTER)'.
 VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
 FORMAT filter_$ (f1.0).
 FILTER BY filter_$.
 EXECUTE.
 COXREG MYS1
   /STATUS=censor(1)
   /PATTERN BY x1
   /PATTERN BY x2
   /PATTERN BY x3
   /PATTERN BY x4
   /CONTRAST (x1)=Indicator
   /CONTRAST (x2)=Indicator
   /CONTRAST (x3)=Indicator
   /CONTRAST (x4)=Indicator
   /METHOD=BSTEP(LR) x1 x2 x3 x4 x5 x6_n
   /PLOT SURVIVAL HAZARDS LML

   /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

Untuk Daerah Perdesaan

```
USE ALL.
 COMPUTE filter_$=(x7=2).
 VARIABLE LABEL filter_$ 'x7=2 (FILTER)'.
 VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
 FORMAT filter_$ (f1.0).
 FILTER BY filter_$.
 EXECUTE.
 COXREG MYS1
   /STATUS=censor(1)
   /PATTERN BY x1
   /PATTERN BY x2
   /PATTERN BY x3
   /PATTERN BY x4
   /CONTRAST (x1)=Indicator
   /CONTRAST (x2)=Indicator
   /CONTRAST (x3)=Indicator
   /CONTRAST (x4)=Indicator
   /METHOD=BSTEP(LR) x1 x2 x3 x4 x5 x6_n
   /PLOT SURVIVAL HAZARDS LML

   /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

LAMPIRAN 4. Program Estimasi Distribusi *Mixture* Pendekatan *Bayesian* Menggunakan WinBUGS 1.4

```
model;
{
   for (i in 1:N){
   Lambda[i] <- pLambda[P[i]]
   }
pLambda[1] ~ dgamma(0.02,130)
pGamma[1] ~ dgamma(1.5,130)

for(i in 1:N){
   P[i] ~ dcat(Phi[1:2])
   }
Phi[1:2] ~ ddirch(Alpha[])
for (i in 1:N){
   Gamma[i] <- pGamma[P[i]]
   }
for (i in 1:N){
   t[i] ~ dweib(Gamma[i],Lambda[i])
   }
pGamma[2] ~ dgamma(2.0,130)
pLambda[2] ~ dgamma(0.01,130)
}
Inits
list(pLambda=c(1,1),pGamma=c(1,1))


Data
list(t=c(16,16,16,16, …, 0.000000001,0.000000001,0.000000001,0.000000001),
N=1394,Alpha=c(1,1),
P=c(1,1,1,1, …, 2,2,1,1))
```

LAMPIRAN 5. Program Model *Mixture* Weibull *Proportional Hazard*

```
model;
{
        for (i in 1:N){
        P[i] ~ dcat(Phi[1:2])
        }
        for (i in 1:N){
        Gamma[i] <- pGamma[P[i]]
        }

for (i in 1:N){
        Lambda[i]<- exp(b1[P[i]]*x1[i] + b2[P[i]]*x2[i] + b3[P[i]]*x3[i] + b4_1[P[i]]*x4_1[i] +
b4_2[P[i]]*x4_2[i]+ b4_3[P[i]]*x4_3[i] +  b5[P[i]]*x5[i] + b6[P[i]]*x6[i])
        }
for (i in 1:N){
t[i] ~ dweib(Gamma[i],Lambda[i])I(t.cen[i],)
}

Phi[1:2] ~ ddirch(Alpha[])
pGamma[1] ~ dgamma(2.009,40)
pGamma[2] ~ dgamma(1.095,50)

tau[1] ~ dgamma(1,150)
sigma[1]<- 1/sqrt(tau[1])
b1[1] ~ dnorm( 0.163,151)
b2[1] ~ dnorm( -0.36,150)
b3[1] ~ dnorm( -0.276,138)
b4_1[1] ~ dnorm( 0.235,230)
b4_2[1] ~ dnorm( 0.217,140)
b4_3[1] ~ dnorm( 0.245,140)
b5[1] ~ dnorm( 0.011,300)
b6[1] ~ dnorm( -0.011,180)

tau[2] ~ dgamma(1,150)
sigma[2]<- 1/sqrt(tau[2])
b1[2] ~ dnorm( 0.045,151)
b2[2] ~ dnorm( -0.228,150)
b3[2] ~ dnorm( 0.294,138)
b4_1[2] ~ dnorm( 0.428,250)
b4_2[2] ~ dnorm( 0.327,140)
b4_3[2] ~ dnorm( 0.120,140)
b5[2] ~ dnorm( -0.005,300)
b6[2] ~ dnorm( -0.156,180)
}

Initialisisasi
list(P(b1=c(0,0),b2=c(0,0),b4_1=c(0,0),b4_2=c(0,0),b4_3=c(0,0),b3=c(0,0),
b5=c(0,0),b6=c(0,0),pGamma=c(1,1),tau=c(1,1))

Data
list(t=c(16,16,16,16, …,9, 9, 9, NA, NA, …, NA),
t.cen=c(0,0,0,…, 0.000000001,0.000000001,0.000000001),
x1=c(1,0,0,1, …, 0,0,1,0),
x2=c(0,1,0,0, …, 1,0,0,0),
.
.
.
x6=c(1.0090746,0.399131,1.502254,0.5200857, …, ,0.6076321,0.2780952,0.3538889,0.4714429),
N=1394,Alpha=c(1,1),P=c(1,1,1,1, …, ,2,2,1,1))
```

LAMPIRAN 6. *Output* Regresi Cox *Proportional Hazard* dengan SPSS 16

Untuk Daerah Perkotaan

Omnibus Tests of Model Coefficients

-2 Log Likelihood
5494.300

Omnibus Tests of Model Coefficients[e,f]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	5474.594	19.306	8	.013	19.706	8	.012	19.706	8	.012
2[b]	5474.616	19.288	7	.007	.022	1	.882	19.684	7	.006
3[c]	5475.056	18.750	6	.005	.440	1	.507	19.244	6	.004
4[d]	5476.938	16.926	5	.005	1.882	1	.170	17.362	5	.004

a. Variable(s) Entered at Step Number 1: x1 x2 x3 x4 x5 x6

b. Variable Removed at Step Number 2: x6

c. Variable Removed at Step Number 3: x5

d. Variable Removed at Step Number 4: x3

e. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 5494.300

f. Beginning Block Number 1. Method = Backward Stepwise (Likelihood Ratio)

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	x1	.163	.091	3.176	1	.075	1.177
	x2	-.360	.104	11.904	1	.001	.698
	x3	-.276	.218	1.604	1	.205	.759
	x4			6.071	3	.108	
	x4(1)	.235	.161	2.136	1	.144	1.264
	x4(2)	.217	.123	3.100	1	.078	1.242
	x4(3)	.245	.125	3.832	1	.050	1.277
	x5	.011	.020	.302	1	.583	1.011
	x6	-.001	.007	.022	1	.882	.999
Step 2	x1	.162	.091	3.157	1	.076	1.176
	x2	-.360	.104	11.949	1	.001	.698
	x3	-.275	.218	1.589	1	.207	.760
	x4			6.728	3	.081	
	x4(1)	.239	.158	2.305	1	.129	1.270
	x4(2)	.220	.121	3.340	1	.068	1.246
	x4(3)	.248	.123	4.059	1	.044	1.281
	x5	.012	.018	.445	1	.504	1.012
Step 3	x1	.163	.091	3.199	1	.074	1.177
	x2	-.359	.104	11.834	1	.001	.699
	x3	-.304	.213	2.035	1	.154	.738
	x4			6.584	3	.086	
	x4(1)	.247	.157	2.469	1	.116	1.280
	x4(2)	.209	.119	3.074	1	.080	1.233
	x4(3)	.246	.123	4.011	1	.045	1.279
Step 4	x1	.150	.091	2.751	1	.097	1.162
	x2	-.340	.103	10.885	1	.001	.711
	x4			6.773	3	.079	
	x4(1)	.278	.155	3.226	1	.072	1.321
	x4(2)	.205	.119	2.942	1	.086	1.227
	x4(3)	.238	.123	3.752	1	.053	1.269

LAMPIRAN 6. *Output* Regresi Cox *Proportional Hazard* dengan SPSS 16 (lanjutan)

<u>Untuk Daerah Perdesaan</u>

Omnibus Tests of Model Coefficients

-2 Log Likelihood
6578.263

Omnibus Tests of Model Coefficients[e,f]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	6551.539	26.080	8	.001	26.724	8	.001	26.724	8	.001
2[b]	6551.593	26.038	7	.000	.054	1	.817	26.670	7	.000
3[c]	6551.899	25.745	6	.000	.305	1	.581	26.365	6	.000
4[d]	6553.920	23.955	5	.000	2.022	1	.155	24.343	5	.000

a. Variable(s) Entered at Step Number 1: x1 x2 x3 x4 x5 x6

b. Variable Removed at Step Number 2: x5

c. Variable Removed at Step Number 3: x1

d. Variable Removed at Step Number 4: x6

e. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 6578.263

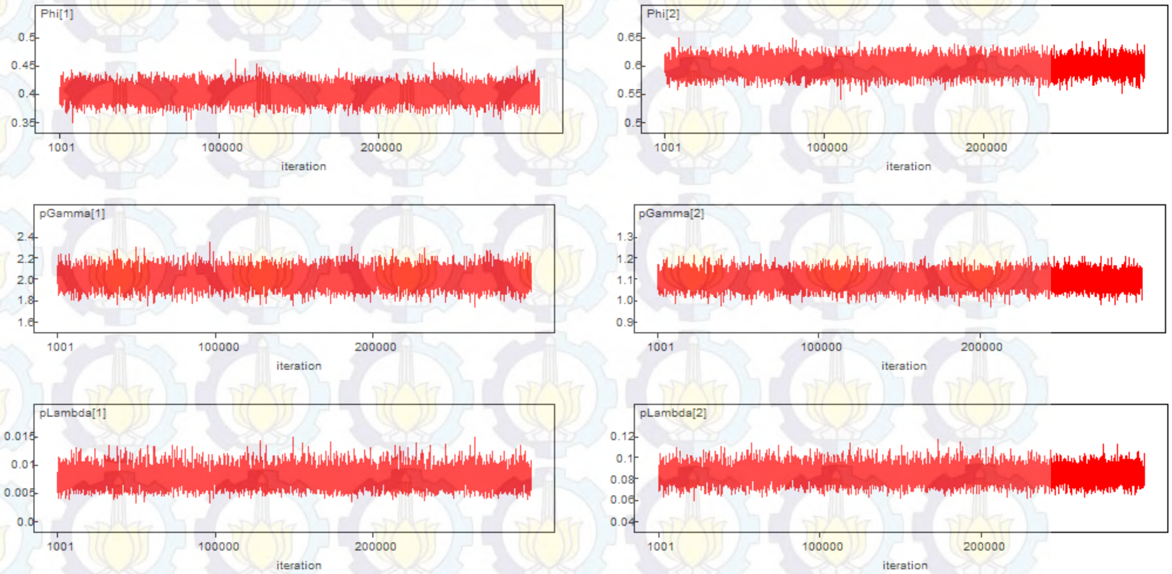
f. Beginning Block Number 1. Method = Backward Stepwise (Likelihood Ratio)

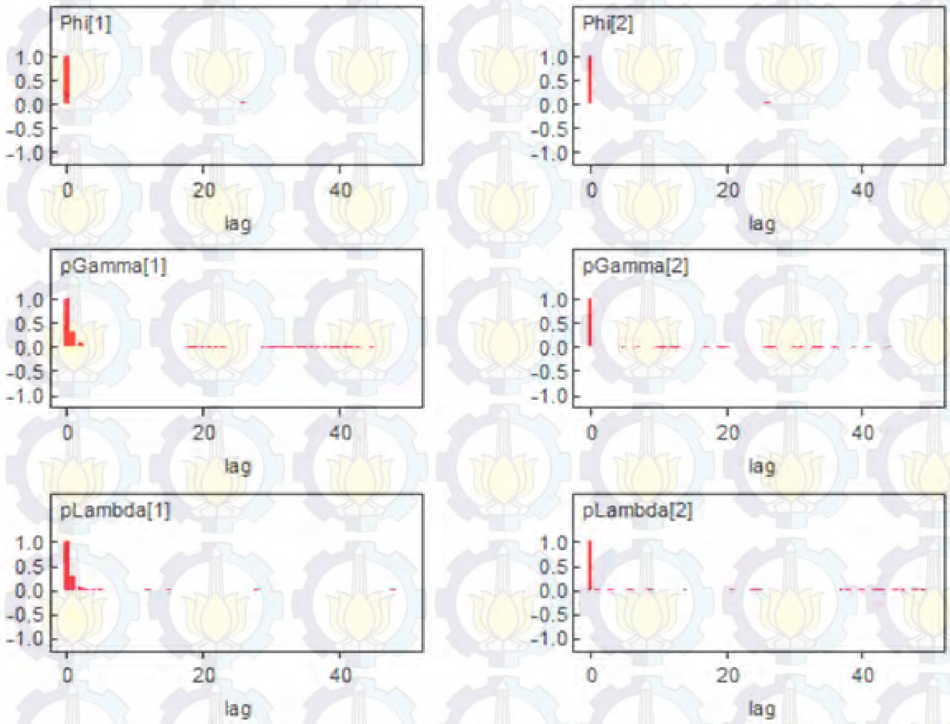
Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	x1	.045	.085	.287	1	.592	1.046
	x2	-.228	.094	5.936	1	.015	.796
	x3	.294	.165	3.167	1	.075	1.342
	x4			14.421	3	.002	
	x4(1)	.428	.128	11.209	1	.001	1.534
	x4(2)	.327	.111	8.757	1	.003	1.387
	x4(3)	.120	.132	.835	1	.361	1.128
	x5	-.005	.020	.054	1	.817	.995
	x6	-.016	.012	1.795	1	.180	.985
Step 2	x1	.047	.085	.305	1	.581	1.048
	x2	-.225	.093	5.884	1	.015	.798
	x3	.299	.164	3.320	1	.068	1.349
	x4			14.656	3	.002	
	x4(1)	.430	.127	11.389	1	.001	1.537
	x4(2)	.330	.110	8.930	1	.003	1.390
	x4(3)	.122	.131	.867	1	.352	1.130
	x6	-.015	.011	1.797	1	.180	.985
Step 3	x2	-.222	.093	5.745	1	.017	.801
	x3	.309	.163	3.591	1	.058	1.362
	x4			14.608	3	.002	
	x4(1)	.431	.127	11.457	1	.001	1.539
	x4(2)	.327	.110	8.830	1	.003	1.387
	x4(3)	.124	.131	.889	1	.346	1.132
	x6	-.015	.011	1.904	1	.168	.985
Step 4	x2	-.223	.093	5.777	1	.016	.800
	x3	.314	.163	3.692	1	.055	1.368
	x4			18.096	3	.000	
	x4(1)	.471	.125	14.334	1	.000	1.602
	x4(2)	.363	.107	11.374	1	.001	1.437
	x4(3)	.163	.129	1.617	1	.204	1.178

LAMPIRAN 7. *Output* Estimasi Parameter Distribusi *Mixture* dengan WinBUGS 1.4

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Phi[1]	0.4012	0.01319	1.282E-4	0.3752	0.401	0.4271	1001	11960
Phi[2]	0.5988	0.01319	1.282E-4	0.5729	0.599	0.6248	1001	11960
pGamma[1]	2.009	0.07723	9.121E-4	1.863	2.007	2.164	1001	11960
pGamma[2]	1.095	0.03486	3.132E-4	1.029	1.094	1.165	1001	11960
pLambda[1]	0.007651	0.001498	1.736E-5	0.005069	0.007535	0.01084	1001	11960
pLambda[2]	0.08406	0.007372	6.875E-5	0.07047	0.08377	0.09908	1001	11960

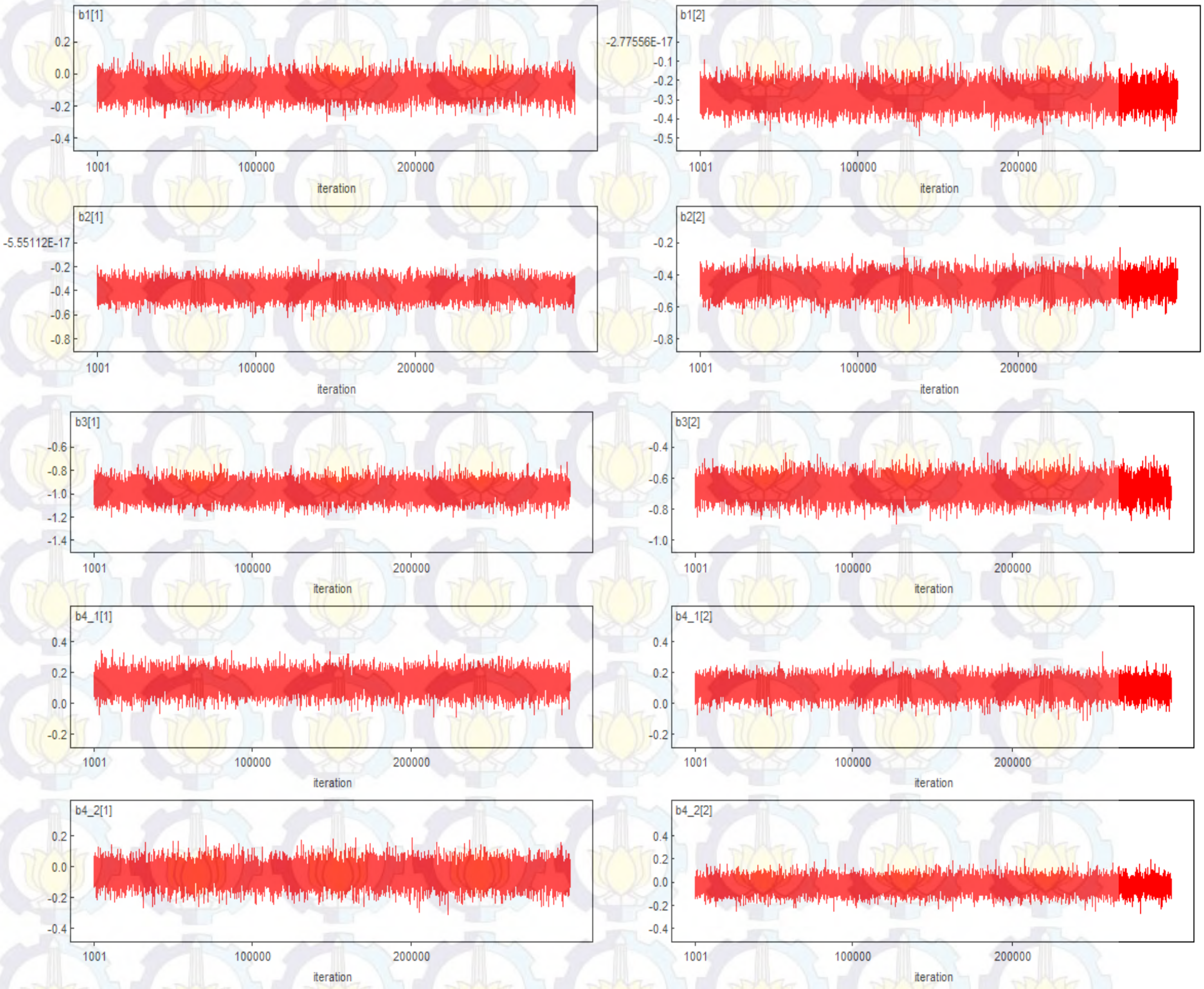


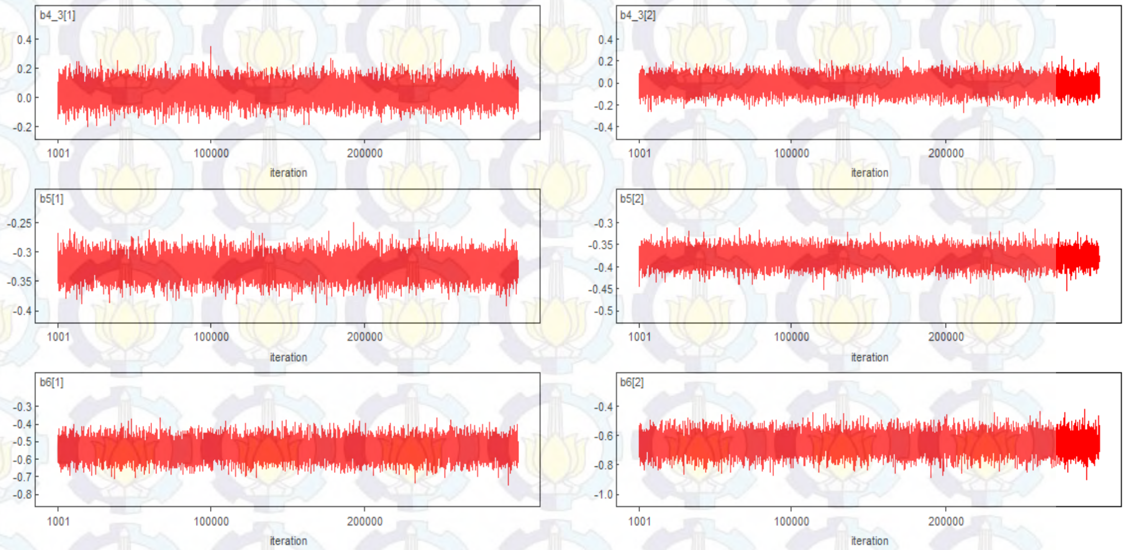


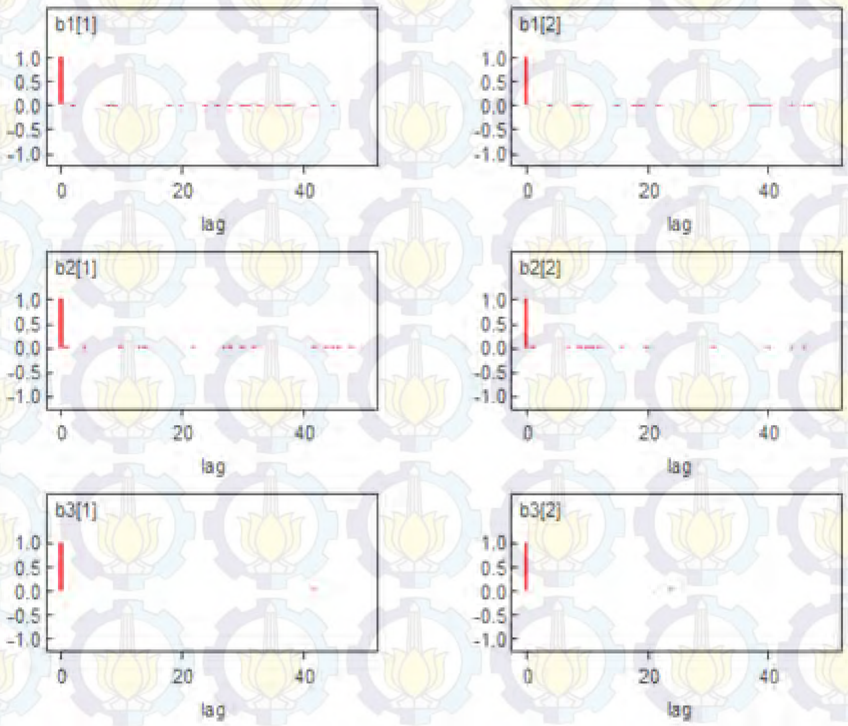
Waktu *running* : 914 s

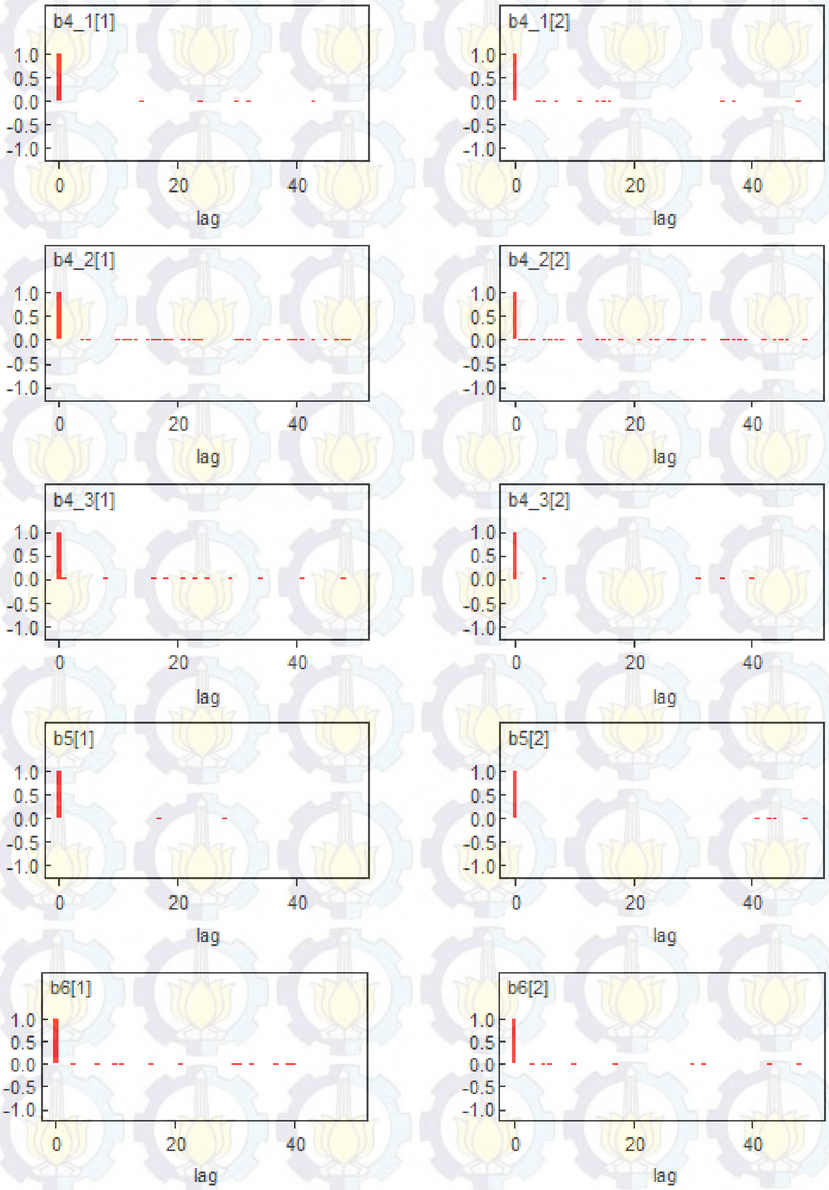
LAMPIRAN 8. *Output* Model *Mixture* Weibull *Proportional Hazard* dengan WinBUGS 1.4

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b1[1]	-0.08322	0.05705	4.882E-4	-0.1949	-0.0834	0.02725	1001	11960
b1[2]	-0.2816	0.05347	5.025E-4	-0.386	-0.2812	-0.1764	1001	11960
b2[1]	-0.3921	0.06243	5.729E-4	-0.5158	-0.3918	-0.2701	1001	11960
b2[2]	-0.4617	0.05679	5.727E-4	-0.5722	-0.4618	-0.3496	1001	11960
b3[1]	-0.9758	0.06928	5.89E-4	-1.111	-0.9759	-0.8386	1001	11960
b3[2]	-0.6626	0.06352	6.452E-4	-0.7864	-0.6625	-0.5378	1001	11960
b4_1[1]	0.1348	0.05865	5.531E-4	0.01855	0.1347	0.2481	1001	11960
b4_1[2]	0.1011	0.05148	5.239E-4	1.677E-4	0.1012	0.2016	1001	11960
b4_2[1]	-0.05012	0.06373	5.98E-4	-0.1734	-0.05081	0.0756	1001	11960
b4_2[2]	-0.02872	0.0584	5.036E-4	-0.1424	-0.02829	0.08651	1001	11960
b4_3[1]	0.03851	0.06569	5.808E-4	-0.08983	0.03848	0.1676	1001	11960
b4_3[2]	-0.01243	0.06664	7.078E-4	-0.144	-0.01257	0.1194	1001	11960
b5[1]	-0.3251	0.01773	1.59E-4	-0.3594	-0.3252	-0.2901	1001	11960
b5[2]	-0.3783	0.0171	1.505E-4	-0.4125	-0.3781	-0.3453	1001	11960
b6[1]	-0.5411	0.05005	3.97E-4	-0.6397	-0.5405	-0.4431	1001	11960
b6[2]	-0.6546	0.06101	5.04E-4	-0.774	-0.6545	-0.533	1001	11960




b4_3[1]
b4_3[2]
b5[1]
b5[2]
b6[1]
b6[2]
1001
100000
200000
iteration


b1[1]
b1[2]
b2[1]
b2[2]
b3[1]
b3[2]
lag



Waktu *running* : 9.245 s

LAMPIRAN 9. Tabel Analisis Deskriptif Karakteristik Penduduk Usia 16 – 24 Tahun dengan Lama Sekolah < 14 Tahun dan ≥ 14 Tahun

a. Persentase Lama Sekolah untuk Variabel Prediktor Kategori

Variabel	Kategori	Persentase Lama Sekolah < 14 Tahun		Persentase Lama Sekolah ≥ 14 Tahun	
		Perkotaan	Perdesaan	Perkotaan	Perdesaan
Jenis Kelamin (X1)	Laki-laki	56,02	57,81	50,00	50,75
	Perempuan	43,98	42,19	50,00	49,25
Status Bekerja (X2)	Bekerja	26,26	40,10	32,35	26,87
	Tidak Bekerja	73,74	59,90	67,65	73,13
Status Perkawinan (X3)	Tidak Kawin	93,87	91,67	96,08	86,57
	Kawin	6,13	8,33	3,92	13,43
Pendidikan KRT (X4)	Tidak Tamat SD	12,69	29,17	8,82	10,45
	Tamat SD	22,10	34,11	15,69	23,88
	Tamat SMP	17,94	13,80	16,67	16,42
	Tamat ≥ SMA	47,26	22,92	58,82	49,25

b. Rata-rata Jumlah ART dan Pengeluaran Perkapita per bulan

Variabel	Lama Sekolah < 14 Tahun		Lama Sekolah ≥ 14 Tahun	
	Perkotaan	Perdesaan	Perkotaan	Perdesaan
Jumlah ART (X5)	5,90	5,93	5,86	5,79
Rata-rata Pengeluaran Perkapita /bulan dlm juta rupiah (X6)	1,045	0,543	1,247	0,723

halaman ini sengaja dikosongkan

BIOGRAFI PENULIS

Penulis dilahirkan di Malang, Jawa Timur pada tanggal 16 Januari 1986, adalah anak pertama dari dua bersaudara dari pasangan Drs. H. Suwari dan Ibu Nur Hamami. Penulis menyelesaikan pendidikan SD hingga Perguruan Tinggi (S1) di Malang dengan jurusan Statistika di Universitas Brawijaya Malang pada tahun 2003. Setelah lulus S1 tahun 2007 penulis aktif bekerja sebagai tenaga pengajar les privat di Malang lalu diterima bekerja sebagai staf *Merchandiser* salah satu perusahaan *garment* di Surabaya pada Februari 2008 sampai Maret 2009. Sembari bekerja, penulis mengikuti seleksi CPNS Badan Pusat Statistik (BPS) 2008 dan diterima di BPS. Pada tahun 2009, penulis ditempatkan di Kabupaten Jayawijaya Provinsi Papua sebagai staf seksi Integrasi Pengolahan dan Diseminasi Statistik (IPDS). Selanjutnya, tahun 2011 penulis mutasi ke Kabupaten Manokwari Provinsi Papua Barat sebagai staf seksi produksi dan diangkat sebagai Kepala Seksi IPDS pada tahun 2013.

Alhamdulillah, pada bulan Juli 2014 penulis kembali diberikan kesempatan untuk melanjutkan pendidikan S2 di Institut Teknologi Sepuluh Nopember (ITS) Surabaya Jurusan Statistika dengan mendapatkan beasiswa APBN-BPS. Di samping karena latar pendidikan sebelumnya statistik, penulis selalu terdorong untuk mengenal suatu metode yang ilmiah dan baru dalam penelitian statistik. Motto dalam setiap langkah kehidupan penulis yaitu "Berusahalah dan berdoalah sekuatnya, karena Allah tidak pernah tidur", sebuah motto yang selalu memotivasi penulis dalam menyelesaikan studi dengan sebaik-baiknya. Pengalaman yang diperoleh penulis selama kuliah S2 Statistika di ITS akhirnya memberikan pelajaran dan pengalaman hidup yang sangat berharga sebagai proses pendewasaan diri dan rasa syukur yang luar biasa. Semoga penulis dapat mengamalkan ilmu yang telah didapat serta mengimplementasikan dalam dunia kerja sekaligus menjadi amal ibadah yang akan dicatat Allah SWT. Amiin.

maulidiah@bps.go.id
hasna.alfaiz@gmail.com