Pemodelan Angka Lama Sekolah Di Provinsi Papua Barat dengan Pendekatan Model *Mixture Survival Bayesian*

[1]Maulidiah Nitivijaya, [2]Nur Iriawan, [3]Heri Kuswanto
[1,2,3]Statistika FMIPA, Institut Teknologi Sepuluh November (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
email: maulidiah@bps.go.id, nur_i@statistika.its.ac.id, kuswanto.its@gmail.com

***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*— lama sekolah, *mixture survival*, MCMC, regresi Cox.**

I. PENDAHULUAN

Salah satu upaya membangun sebuah bangsa yang maju dan modernya itu melalui pendidikan. Selain sebagai sarana transfer pengetahuan dan keterampilan, pendidikan semestinya menjadi proses pembelajaran sepanjang hidup untuk membentuk karakter yang baik, mengembangkan potensi, talenta individu serta serta jiwa yang mandiri. Ketika pendidikan semakin tinggi maka kebudayaan dan peradaban yang unggul dan maju akan terbentuk. Seiring dengan yang tertuang pada RPJMN 2015-2019 yang salah satu agenda utama dalam bidang pendidikan adalah 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. Namun pelaksanaan program tersebut mencakup keseluruhan proses pendidikan hingga siswa mampu menyelesaikan jenjang pendidikan menengah, maka berbagai permasalahan dalam pelaksanaan Wajar 9 Tahun yang belum terselesaikan harus dapat diatasi.

Pendidikan merupakan salah satu pilar dalam menyusun Indeks Pembangunan Manusia (IPM). Pada tahun 2013, indeks pendidikan pada penghitungan IPM di Indonesia masih diwakili oleh dua indikator yaitu Angka Melek Huruf (AMH) dan rata-rata lama bersekolah (*mean years of schooling*). AMH 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 yang 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 AMH tidak lagi digunakan. Cakupan indikatorpendidikan 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) 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.

Memperhatikan adanya target program peningkatan tingkat pendidikan, maka perlu dikaji tentang variabel-variabel apa yang mempengaruhi lamanya sekolah penduduk, 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. 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.

Dalam statistika terdapat metode analisis *survival* dimana metode tersebut mempelajari lamanya suatu peristiwa atau kejadian yang biasa disebut *failure event*. Waktu dari awal perlakuan sampai terjadinya respon pertama kali yang ingin diamati disebut sebagai waktu ketahanan hidup (*survival time*). 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, seringkali ditemui adanya pola distribusi data yang tidak mudah diamati. 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*. 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). 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.

Provinsi Papua Barat yang terbagi atas 11 kabupaten/kota memiliki wilayah perkotaan dan perdesaan dimana wilayah perdesaan lebih dominan dibandingkan perkotaan. Dalam kenyataannya, 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. 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.

Tujuan penelitian ini yaitu untuk memperoleh cara estimasi parameter dan model lama sekolah dengan pendekatan model *mixture survival Bayesian* berdasarkan klasifikasi daerah serta mendapatkan variabel-variabel yang berpengaruh terhadap lama sekolah di Provinsi Papua Barat menggunakan model *mixture survival Bayesian* berdasarkan metode *cox proportional hazard*.

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.

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

II. TINJAUAN PUSTAKA

A. Pengujian Distribusi Data

Dalam analisis *survival* yang menjadi respon adalah data waktu dari suatu objek hingga terjadinya suatu kejadian tertentu. Pengujian distribusi data variabel respon *(goodness of fit)* dapat dilakukan dengan beberapa cara, diantaranya dengan menggunakan metode uji Anderson Darling, Kolmogorov-Smirnov, dan Chi-Square.

Statistik uji *goodness of fit* dengan metode Anderson Darling yaitu sebagai berikut:

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

dimana F merupakan fungsi distribusi kumulatif dari distribusi yang dihipotesiskan dan x_i merupakan data waktu *survival* yang telah diurutkan. Dengan hipotesis

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

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

Pengambilan keputusan tolak H_0 apabila $A_n^2 > a_{n,1-\alpha}$ atau *p-value*$< \alpha$, 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).

B. Analisis Survival

Analisis *survival* merupakan 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, sedangkan 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.

Terdapat tiga elemen yang harus diperhatikan dalam menentukan waktu *survival* (Zang, 2008), yaitu: *time origin* atau waktu awal dimulai penelitian, *failure time,* dan skala pengukuran waktu atau *measurement scale of time.* Oleh karenanya dalam analisis survival memungkinkan adanya objek yang tidak teramati secara penuh hingga terjadi failure event atau disebut dengan data tersensor. Collet (1994) berpendapat, terdapat tiga alasan terjadinya penyensoran, diantaranya adalah: *lost to follow up* yaitu jika obyek pengamatan meninggal, pindah, atau menolak untuk berpartisipasi; *drop out* yaitu jika perlakuan harus dihentikan karena suatu alasan tertentu; dan *termination of study* yaitu jika masa penelitian berakhir sementara obyek pengamatan belum mencapai pada *failure event.*

C. Fungsi Survival dan Fungsi Hazard

Dalam analisis survival, waktu survival digambarkan dalam tiga fungsi yaitu fungsi kepadatan peluang (*probability density function*), fungsi *survival*, dan fungsi *hazard.* Fungsi *survival* $S(t)$ dapat dinyatakan sebagai peluang seseorang dapat bertahan lebih lama dari suatu waktu t dan dinyatakan melalui persamaan berikut

$$S(t) = \Pr(T \geq t) = 1 - \Pr(T < t) = 1 - F(t). \qquad (2)$$

Fungsi *hazard* atau $h(t)$ merupakan reaksi sesaat atau laju kegagalan (*failure*) sesaat ketika seseorang mengalami suatu *event* pada waktu ke-t. Hubungan antara fungsi *survival*, fungsi *hazard,* dan fungsi kepadatan peluang waktu survival yaitu sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)}, \qquad (3)$$

sedangkan hubungan antara fungsi kumulatif *hazard* yang dilambangkan $H(t)$ dengan fungsi *survival* yang dilambangkan $S(t)$ adalah

$$H(t) = -\ln S(t). \qquad (4)$$

D. Pemodelan Fungsi Hazard

Model umum proportional hazard atau dikenal sebagai regresi Cox adalah

$$h_i(t) = h_0(t)\exp\left(\beta_1 x_{1i} + \ldots + \beta_p x_{pi}\right). \qquad (5)$$

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 (x=1) dengan kategori gagal (x=0) (Hosmer dan Lemeshow, 1999) atau dapat dituliskan

$$\frac{h(t|x=1)}{h(t|x=0)} = \frac{h_0(t)e^{\beta.1}}{h_0(t)e^{\beta.0}} = \frac{h_0(t)e^{\beta}}{h_0(t)} = e^{\beta}, \qquad (6)$$

yang artinya 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

E. Asumsi Hazard Proporsional

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 membentuk pola yang sejajar (Kleinbaum dan Klein, 2005).

F. Estimasi Parameter Regresi Cox Proportional Hazard

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

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}'\mathbf{x}_l)}, \qquad (7)$$

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 (7) 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:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{x}_l)} \right]^{\delta_i}, \qquad (8)$$

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

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left(\boldsymbol{\beta}'\mathbf{x}_i - \ln \sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{x}_l) \right). \qquad (9)$$

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.

Collet (2003) menyatakan bahwa estimasi dengan metode *partial likelihood* yang ditunjukkan pada persamaan (8) dalam prakteknya sulit untuk diterapkan karena asumsi *partial likelihood* hanya berlaku jika data penelitiannya seluruhnya sudah mengalami kondisi *failure.* Akan tetapi pada kenyataannya dalam penelitiandata *survival* selalu memiliki keterbatasan dengan adanya data tersensor, sehingga penghitungan komputasi *partial likelihood* dilakukan dengan pendekatan metode Breslow yang dituliskan sebagai berikut:

$$\prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}' s_{(j)})}{\left\{\sum_{l \in R(t_{ij})} \exp(\boldsymbol{\beta}' \mathbf{x}_{(l)})\right\}^{d_j}}, \quad (10)$$

dengan s_j menunjukkan banyaknya variabel prediktor $\mathbf{x}_{(j)}$yang mengalami kondisi *failure* pada waktu $t_{(j)}$ (Breslow, 1974 dalam Collet 1994).

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:

$$G^2 = -2\ln\left[\frac{L_0}{L_p}\right], \quad (11)$$

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^2ini 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 regresi 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:

$$W_k = \left[\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}\right]^2 \quad (12)$$

dengan keterangan:

$\hat{\beta}_k$ merupakan penduga β_k

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

Statistik $W_k > \chi^2_{\alpha;1}$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.

G.Model Mixture

Model *mixture* merupakan suatu model yang khas dimana terlihat dari data yang diamati dan data yang adaterdiri 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).

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 :

$$f(t|\pi) = \pi_1 f_1(t) + \pi_2 f_2(t) + \cdots + \pi_M f_M(t). \quad (13)$$

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 (13) disebut sebagai model *finite mixture* yang berlaku untuk model dengan jumlah komponen *M* tertentu.

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). 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).

H.Analisis Bayesian

Dalam pendekatan *Bayesian*, data sampel yang diperoleh dari populasi, juga memperhitungkan suatu distribusi awal yang disebut *prior*. 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* ∝ *likelihood* × Distribusi *prior*

Pada teorema Bayes, apabila terdapat parameter θ yang diberikan oleh data observasi *survival timeT*, 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:

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

atau

$$f(\theta|t) \propto f(t|\theta)f(\theta). \quad (14)$$

Menurut Iriawan (2001), jika diterapkan pada model *mixture* maka θ dapat disinkronisasi sebagai sebuah parameter yang memuat semua parameter model penyusun *mixture*.

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 titikapakah terdistribusi secara *uniform* atau tidak (Ntzoufras, 2009).
3. *Informative prior* atau *non informative prior* yaitupenentuan *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*.

I. Fungsi Likelihood pada Model Mixture

Menurut McLachlan dan Basford (1988), fungsi *likelihood* pada distribusi *mixture* berbeda dengan fungsi *likelihood* pada distribusi univariat biasa.

Berdasarkan model *mixture* pada persamaan (13), fungsi *likelihood* model *mixture* adalah:

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

$$l_{mix} = \prod_{i_1=1}^{n_1} f(t_{i1}|\pi_1, \theta_1) + \prod_{i_2=1}^{n_2} f(t_{i2}|\pi_2, \theta_2) + \ldots + \prod_{i_m=1}^{n_M} f(t_{iM}|\pi_M, \theta_M), \quad (15)$$

dengan syarat persamaan (15) adalah $n_1+n_2+ \ldots + n_M = n$ dan M adalah banyaknya komponen *mixture*.

J. 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).

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). 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*.

K. 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 (13), dapat dibentuk model *mixture* regresi *survival* denganfungsi 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:

$$p(t|\pi, \theta) = \sum_{m=1}^{M} \pi_m p(t|\theta_m). \quad (16)$$

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:

$$p(t|\pi, \theta) = \pi_1 p(t|\theta_1) + \pi_2 p(t|\theta_2), \quad (17)$$

dimana

$p(t|\theta_1)$: fungsi densitas untuk data *suvival* daerah perkotaan

$p(t|\theta_2)$: fungsi densitas untuk data *suvival* daerah perdesaan

π_1 : kontribusi untuk komponen distribusi *mixture* daerah perkotaan

$\pi_2 = (1 - \pi_1)$: kontribusi untuk komponen distribusi *mixture* daerah perdesaan

L. 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.

M. Angka Lama Sekolah dan Rata-Rata Lama Sekolah

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 (*under estimate*).

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:

$$ALS = \text{Tahun\_Konversi} + (\text{Kelas\_Tertinggi} - 1). \quad (18)$$

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 = 12 tahun; D3 = 13 tahun; D4/S1 = 16 tahun; S2 = 18 tahun; S3 = 21 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.

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.

III. METODOLOGI PENELITIAN

A. 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.

B. Variabel Penelitan

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:

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.

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: masih bersekolah pada jenjang SD/sederajat atau SMP/sederajat;jika seorang responden tidak pernah sama sekali bersekolah;jika seorang responden keluar dari masa belajar (mengundurkan diri) atau mengalami *drop out* (dikeluarkan dari sekolah), maka data *survival* tersebut dikatakan data tersensor.

Adapun variabel prediktor yang digunakan pada penelitian ini antara lain:

Tabel 1. Variabel Penelitian

Variabel	Deskripsi	Skala	Kategori
X_1	Jenis Kelamin Responden	nominal	1= laki-laki 2 = perempuan
X_2	Status Bekerja	nominal	1= tidak bekerja 2 = bekerja
X_3	Status Perkawinan Responden	nominal	1= tidak kawin 2 = kawin
X_4	Tingkat Pendidikan Kepala Rumah Tangga	ordinal	1= tidak tamat SD 2= tamat SD 3= tamat SMP 4= tamat SMA ke atas
X_5	Jumlah ART	rasio	-
X_6	Rata-rata pengeluaran rumah tangga perkapita sebulan (juta)	rasio	-

C. Tahapan Penelitian

Sebelum melakukan tahapan penelitian, terlebih dahulu dilakukan tahap *pre-processing* data yang akan diolah. Selanjutnya tahapan penelitian yang 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 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 perdesaandan 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. 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.

IV. HASIL DAN PEMBAHASAN

Tabel 2. Deskriptif Penduduk Usia 16 – 24 Tahun Daerah Perkotaan

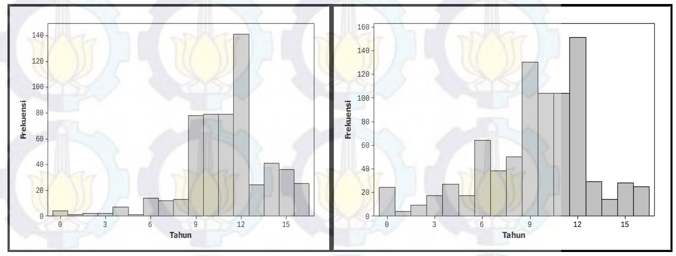
Variabel	Minimum	Maksimum	Mean	Std. Deviasi
Lama Sekolah Penduduk 16 - 24 tahun	0	16	11,08	2,71
Jumlah ART	2	18	5,89	2,49
Rata-rata Pengeluaran Perkapita Sebulan (juta rupiah)	0,2594	5,9998	1,0637	0,7133

Tabel 3. Deskriptif Penduduk Usia 16 – 24 Tahun Daerah Perdesaan

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

Tabel 2 dan 3 diatas didapatkan informasi bahwa waktu *survival* penduduk atau angka lama sekolah penduduk umur 16-24 tahun baik di perkotaan maupun perdesaan dimana paling sedikit 0 tahun yang artinya belum pernah memperoleh pendidikan formal dan paling besar yaitu 16 tahun yang artinya sudah mendapatkan ijazah pada jenjang SMA. Rata-rata lama sekolah penduduk perkotaan yang berusia 16-24 tahun sebesar 11,08 yang artinya sebagian besar penduduk yang berusia 16-24 tahun memiliki rata-rata lama sekolah di atas 9 tahun atau dikatakan sudah lulus pendidikan pada jenjang SMP/sederajat. Sedangkan di perdesaan rata-rata lama sekolah 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 18 orang di perkotaan dan 20 di perdesaan . 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,-. Di perdesaan jauh lebih ekstrim perbedaannya dimana minmal rata-rata pengeluaran perkapita perbulan Rp 144.800,- dan maksimal Rp 8.754.800,-.

A. Distribusi Data Lama Sekolah



(a) (b)

Gambar 1. Histogram Lama Sekolah Penduduk Usia 16-24 Tahun di (a) Perkotaan dan (b) Perdesaan

Histogram data pada Gambar 1 menunjukkan pola data lama sekolah di daerah perkotaan maupun perdesaan. Selanjutnya Tabel 4 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. *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* 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 $-\infty$ sampai ∞. Adanya indikasi distribusi *mixture* dapat dipertimbangkan dimana pendekatan yang digunakan adalah distribusi *weibull* untuk kedua wilayah.

B. Model Survival

Berdasarkan plot fungsi ln(-ln $S(t)$) pada masing-masing variabel prediktor di daerah perkotaan dan perdesaan, 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*.

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

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

Di daerah perdesaan juga dibentuk model regresi *cox* angka lama sekolah dengan keseluruhan variabel prediktor sehingga didapatkan model sebagai berikut:

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

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 menunjukkan pada taraf 10 signifikansi persen, model terbaik regresi *Cox proportional hazard* pada daerah perkotaan dan perdesaan masing-masing adalah:

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

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

C. Model Mixture Survival

Model *mixture* regresi *survival* dengan dua komponen didasarkan pada persamaan (13) dimana fungsi densitasnya disusun dari distribusi data *survival* (*t*) sehingga diperoleh bentuk seperti pada persamaan (17) dapat juga dinyatakan sebagai berikut:

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

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:

$$S(t) = \pi\, S_1(t) + (1-\pi)\, S_2(t), \tag{19}$$

sehingga model *proportional hazard* untuk *mixtures urvival* dapat dinyatakan sebagai berikut:

$$h(t) = \pi\, h_1(t) + (1-\pi)\, h_2(t), \tag{20}$$

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

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

$$h(t) = \pi\, \lambda_1 \gamma_1 t^{\gamma_1 - 1} + (1-\pi)\lambda_2 \gamma_2 t^{\gamma_2 - 1} \tag{21}$$

untuk $0 \leq t \leq \infty$.

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 dapat ditulis sebagai berikut:

$$h_i(t) = \pi \lambda_1 \gamma_1 t^{\gamma_1 - 1} \exp(\boldsymbol{\beta}_1' \mathbf{x}_{i1}) + (1-\pi)\lambda_2 \gamma_2 t^{\gamma_2 - 1} \exp(\boldsymbol{\beta}_2' \mathbf{x}_{i2}), \tag{22}$$

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

$$S^*(t) = \pi \exp(-\lambda_1 t^{\gamma_1}) \exp(\boldsymbol{\beta}_1' \mathbf{x}_{i1}) + (1-\pi) \exp(-\lambda_2 t^{\gamma_2}) \exp(\boldsymbol{\beta}_2' \mathbf{x}_{i2}). \tag{23}$$

Estimasi untuk setiap parameter diperoleh melalui metode *Bayesian* dengan menentukan distribusi *prior* terlebih dahulu. Pada penelitian ini, distribusi *prior* yang digunakan yaitu *prior conjugate, pseudo prior*, dan *informative prior*. *Prior conjugate* dari distribusi weibull adalah distribusi gamma yang berasal dari keluarga eksponensial. Sedangkan distribusi normal digunakan sebagai *prior* dari β. *Pseudo prior* ini digunakan untuk menentukan *initial* parameter dari *prior conjugate* dimana nilainya diperoleh dari estimasi dengan metode maksimum *likelihood*. Selanjutnya, semua informasi diatas menghasilkan *informative prior* untuk estimasi parameter di step berikutnya. Bentuk posterior yang digunakan untuk esimasi parameter model *mixture* weibull *proportional hazard* berdasarkan distribusi prior dan likelihood dapat ditulis sebagai berikut:

$$p(\lambda,\gamma,\pi\,|\,t) \propto l(t\,|\,\lambda,\gamma,\pi) \times p(\lambda) \times p(\gamma) \times p(\pi) \tag{24}$$

dengan

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

D.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*. 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 5.

Tabel 5. 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 5 maka distribusi *mixture weibull* untuk data waktu *survival* lama sekolah penduduk dapat dinyatakan secara matematis menjadi sebagai berikut:

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

Selanjutnya fungsi *survival* dan fungsi *hazard* model *mixture weibull* dapat dituliskan:

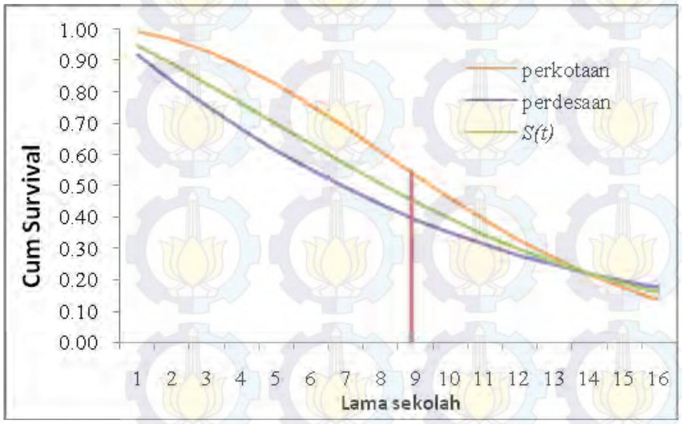
$$\hat{S}(t) = 0,4012\ \hat{S}_1(t) + 0,5988\ \hat{S}_2(t),$$

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.



Gambar 2. Plot Fungsi *Survival* Model *Mixture*

Secara grafis, Gambar 2 menunjukkan plot fungsi *survival* model *mixture* pada 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).

E. Pemodelan Mixture Weibull Proportional Hazard

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 (22) dan dilakukan estimasi parameter model *mixture survival* menggunakan metode *Bayesian* melalui teknik MCMC *gibbs sampling*.

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. Plot yang dihasilkan cenderung memiliki variansi kecil dan berjalan di sekitar rentang nilai tertentu, serta tidak ada yang memiliki jarak yang sangat jauh. Plot autokorelasi juga menunjukkan bahwa data sampel signifikan di lag nol, artinya sudah tidak mengandung autokorelasi sehingga hasil estimasi yang diperoleh dapat dikatakan sudah cukup baik. Hasil estimasi parameter model *mixture weibull proportional hazard* ditunjukkan pada Tabel 6.

Tabel 6. Estimasi Parameter Model *Mixture Weibull Proportional Hazard*

Parameter	*Mean*	*exp(B)*	*Std.Dev*	*2,5%*	*Median*	*97,5%*
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**
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 6 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).

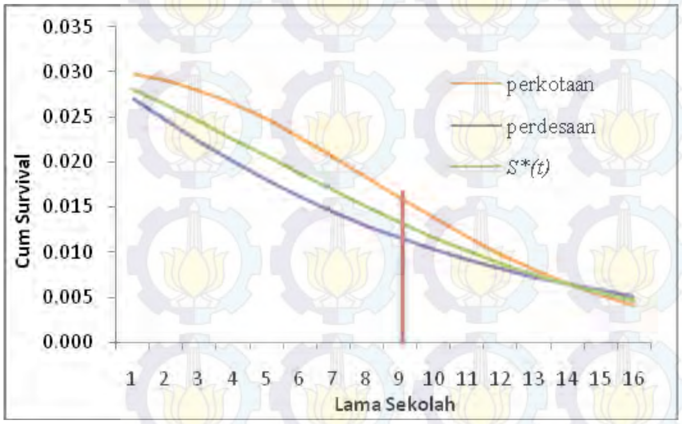
Model hubungan faktor-faktor yang berpengaruh terhadap lama sekolah penduduk usia 16 – 24 tahun di Provinsi Papua Barat berdasarkan persamaan (22) sesuai dengan hasil estimasi parameter *mixture weibull proportional hazard* untuk daerah perkotaan secara matematis dapat dituliskan sebagai berikut:

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

dan model untuk daerah perdesaan adalah:

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

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 (23) dapat disajikan seperti pada Gambar 3.



Gambar 3. Plot Fungsi *Survival* Model *Mixture Weibull Proportional Hazard*

Dari Gambar 3 di atas terlihat 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 menuju ke pendidikan 12 tahun (setingkat SMA). Uniknya pada pendidikan 14 tahun (setara tingkat kedua perguruan tinggi) tidak ada perbedaan antara peluang untuk melanjutkan sekolah antara penduduk di perkotaan dan perdesaan. Hal ini menunjukkan adanya pengaruh yang signifikan dari variabel-variabel dalam model terhadap angka lama sekolah pada masing-masing wilayah.

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.

Interpretasi model *mixture weibull proportional hazard* untuk daerah perkotaan, misalkan untuk variabel status bekerja yaitu ditunjukkan 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.

Sedangkan interpretasi model *mixture weibull proportional hazard* untuk daerah perdesaan, misalkan untuk 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.

V. PENUTUP

A. 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 *mixtureweibull* 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.

B. Saran

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. Oleh karenanya, perlu adanya kebijakan terkait pengaruh pekerja anak terhadap angka lama sekolah baik di daerah perkotaan maupun perdesaan 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. 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.

DAFTAR PUSTAKA

Badan Pusat Statistik (2010), *Peraturan Kepala BPS Nomor 37 Tahun 2010*

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.

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.

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.

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

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.

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.

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.

Muthen, B. dan Masyn, K. (2005), "Discrete-Time *MixtureSurvival* 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.

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.

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

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