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

Oleh:
Maulidiah Nitivijaya
1314 201 709

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

OUTLINE

- PENDAHULUAN
- TINJAUAN PUSTAKA
- METODOLOGI PENELITIAN
- HASIL DAN PEMBAHASAN
- KESIMPULAN DAN SARAN
- DAFTAR PUSTAKA

Bangsa yang maju dan modern

IPM

Indeks Pendidikan

2013

MYS

Usia 15 tahun ke atas

Angka Melek Huruf

2014

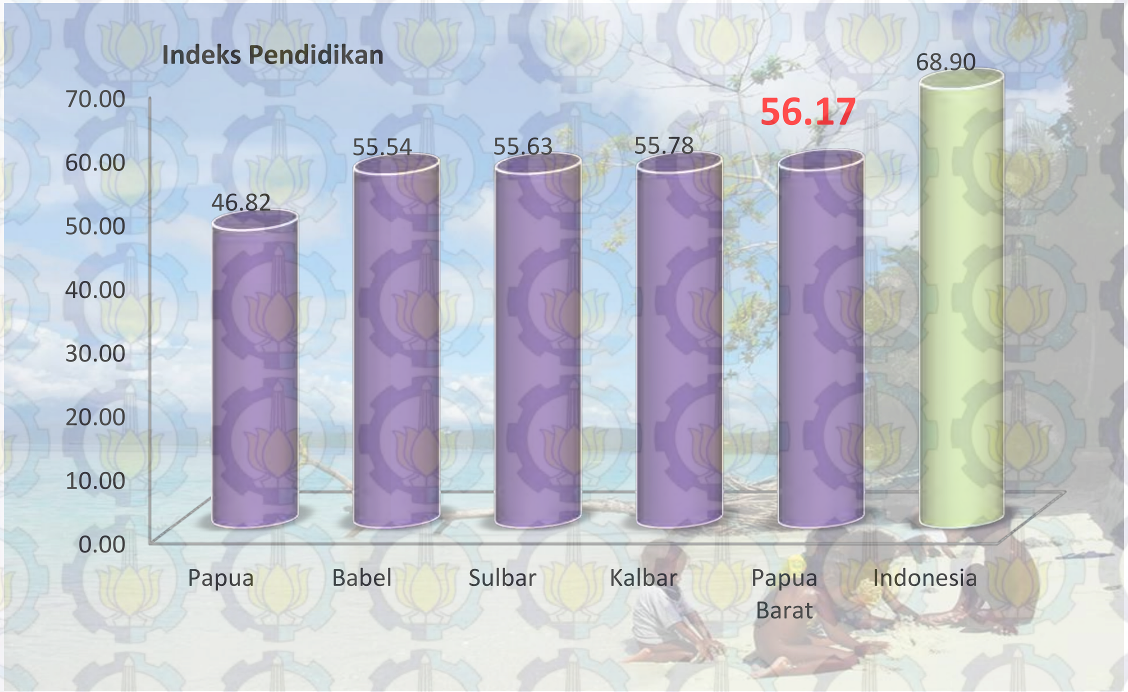
MYS

Usia 25 tahun ke atas

EYS

Wajar 9 Tahun

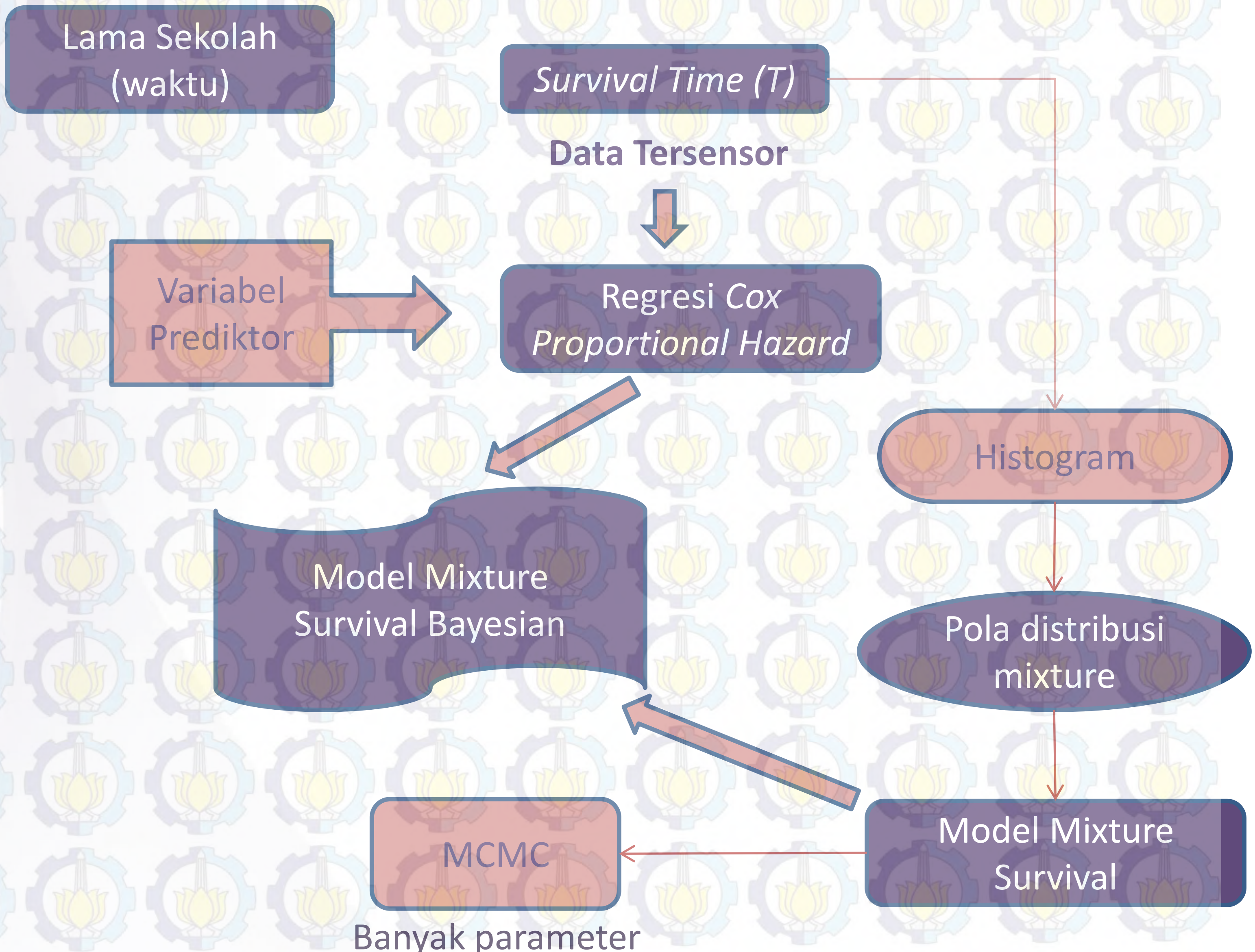
Upaya pengentasan kemiskinan



MYS Papua Barat 2014 = 8,66

Papua Barat tergolong provinsi termuda

Wajar 9 Tahun??


Lama Sekolah (waktu)
Survival Time (T)
Data Tersensor
Variabel Prediktor
Regresi *Cox Proportional Hazard*
Histogram
Model Mixture Survival Bayesian
Pola distribusi mixture
MCMC
Model Mixture Survival
Banyak parameter

Penelitian Sebelumnya

Pendidikan

- **Brunello dan Checchib (2005)**
- **Santoso (2009),** ALS di Provinsi Papua
- **Solikhah (2009)**, MYS di Pulau Kalimantan
- **Setyawan (2011),** AMH dan MYS di Pulau Papua

Mixture Survival

- **Muthen dan Masyn (2005)**
- **Hariyanto (2009)** → lama mencari kerja di Pulau Jawa
- **Qudsi (2015)** → ALS di Provinsi Jatim dengan mixture survival dgn frailty

Bagaimana memodelkan angka lama sekolah anak usia 16-24 tahun di Papua Barat dengan pendekatan model *mixture survival* Bayesian berdasarkan klasifikasi daerah tempat tinggal?

- ✓ **Memperoleh cara estimasi parameter *mixture survival Bayesian* berdasarkan klasifikasi daerah tempat tinggal.**

- ✓ **Memperoleh model lama sekolah dengan pendekatan model *mixture survival* Bayesian berdasarkan klasifikasi daerah tempat tinggal.**

- ✓ **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.**

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

- ✓ Unit observasi dalam penelitian ini adalah penduduk usia 16-24 tahun di Provinsi Papua Barat dengan status anak.
- ✓ Sensor yang digunakan dalam penelitian ini adalah sensor kanan (*right censor*) yang artinya jika responden sampai dengan masa pencacahan selesai belum mengalami *failure event* maka waktunya dibatasi hanya sampai dengan berakhirnya masa pencacahan.

Tahapan awal yang dilakukan yaitu mengetahui distribusi data dari variabel respon melalui uji *goodness of fit*

Beberapa cara, diantaranya dengan menggunakan metode uji Anderson Darling, Kolmogorov-Smirnov, dan Chi-Square

Statistik uji Anderson Darling :

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

Uji Hipotesis

H_0 : data X merupakan variabel random independen yang berdistribusi sesuai dengan distribusi $F\ (x_i)$

H_1 : data X merupakan variabel random independen yang tidak berdistribusi sesuai dengan distribusi $F\ (x_i)$

tolak H_0 apabila $> a_{n,1-\alpha}$ atau p-value $< \alpha$

Data dikatakan mengikuti distribusi tertentu apabila nilai statistik Anderson-Darling semakin kecil (Iriawan dan Astuti, 2006)

Analisis *survival* adalah salah satu metode statistika untuk menganalisis data dimana variabel responnya berupa waktu sampai suatu peristiwa atau *event* terjadi

Event → peristiwa ekstrim yang mungkin terjadi pada individu

Respon Survival → *range* waktu dari awal penelitian sampai suatu *event* terjadi atau sampai penelitian berakhir

Medis | Teknik/industri | Sosiologi | Ekonomi

Tiga elemen yang harus diperhatikan dalam menentukan waktu *survival t* (Zang, 2008)

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
3. *Measurement scale of time*, misal skala waktu yang digunakan dalam penentuan lama sekolah seorang anak

Data Tersensor

Tiga alasan terjadinya penyensoran, diantaranya sebagai berikut (Collet, 1994) :

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
2. Data tersensor kanan
3. Data sensor interval

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

Fungsi *Survival* dan Fungsi *Hazard*

Peluang *T (waktu survival)* → *pdf*, fungsi *survival*, dan fungsi *hazard*

Fungsi distribusi kumulatif : $F(t) = \Pr(T < t) = \int_0^t f(u)\,du \ , 0 < t < \infty$

Fungsi *survival* : $S(t) = \Pr(T \geq t) = 1 - \Pr(T < t) = 1 - F(t)$

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

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

Hubungan antara fungsi *survival* dan fungsi *hazard,* digunakan teori probabilitas bersyarat :

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

dalam bentuk $h(t)$:

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

$$F'(t) = \operatorname*{Lim}_{\Delta t \to 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} = f(t), \quad \Longrightarrow \quad h(t) = \frac{f(t)}{S(t)},$$ dimana $F(t) = 1 - S(t)$

Hubungan antara fungsi kumulatif *hazard* dengan fungsi *survival* :

$$H(t) = -\ln S(t)$$

Pemodelan Fungsi *Hazard*

Fungsi *hazard yaitu* 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 (Lee, 1992)

Fungsi *baseline hazard* dinyatakan sebagai $h_0(t)$, yaitu fungsi *hazard* untuk tiap individu dimana semua variabel prediktor dalam vektor **x** bernilai 0 (Collet, 2003).

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

model umum *proportional hazard* atau Regresi Cox :

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

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

Odds perbandingan antara *Odd* individu dengan kondisi variabel prediktor x pada kategori sukses dengan kategori gagal (Hosmer dan Lemeshow, 1999).

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

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

Asumsi *Hazard* Proporsional

Fungsi *hazard* harus proporsional setiap waktu, karena regresi *cox* tidak mengakomodasi variabel yang berubah-ubah sepanjang waktu Collet (2003)

Plot –ln [–ln $S(t)$] atau ln [–ln $S(t)$] terhadap waktu *survival* untuk setiap variabel prediktor dengan skala kategorik membentuk pola SEJAJAR

Estimasi Parameter Regresi Cox Proportional *Hazard*

Estimasi fungsi *hazard* untuk objek ke-i $\hat{h}_i(t) = e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i}\hat{h}_0(t)$

Koefisien $\boldsymbol{\beta}$ dalam model *proportional hazard* dapat diestimasi dengan metode estimasi *partial likelihood*

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

Jika data terdiri atas n pengamatan waktu *survival* $t_1, t_2, \ldots, t_n$, dan δ_i adalah indikator sensor yang bernilai nol untuk waktu *survival* ke-*i*, t_i (i=1, 2, ..., n) dan tersensor kanan

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

pendekatan metode Breslow

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

Pengujian Parameter Model

Uji Serentak

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$
H_1 : minimal ada satu dengan $k = 1, 2, \ldots, p$

Statistik Uji :

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

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

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

Tolak H_0 jika $G^2 > \chi^2_{\alpha;p}$ atau *p-value* < α

Uji Parsial

$H_0 : \beta_k = 0$
$H_1 : \beta_k \neq 0$ dengan $k = 1, 2, \ldots, p$

Statistik Uji :

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

Tolak H_0 jika $W_k > \chi^2_{\alpha;1}$ atau *p-value* < α

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)

Prinsipnya: menggabungkan sejumlah komponen yang mungkin berasal dari distribusi yang sama atau berbeda-beda sehingga dapat memberikan gambaran mengenai sifat-sifat data.

Model *mixture* yang disusun oleh *k* komponen :

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

dimana $\pi_1 + \pi_2 + \cdots + \pi_M = 1$

Estimasi model *mixture* dengan munculnya banyak parameter menimbulkan model yang kompleks sehingga digunakan pendekatan metode Bayesian melalui proses simulasi *Markov Chain Monte Carlo* (MCMC)

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*

$$f\left(\theta|t\right) \propto f\left(t|\theta\right) f(\theta).$$

Beberapa macam disribusi *prior* (Box dan Tiao, 1973) :

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*.

Fungsi *Likelihood* pada Model *Mixture*

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

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

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

$$n_1 + n_2 + \ldots + n_M = n$$

Markov chain Monte Carlo (MCMC)

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).

Sampel parameter dalam *markov chain* diambil setelah kondisi stasioner tercapai.

Gibbs sampling

Gibbs sampling : suatu teknik simulasi untuk membangkitkan variabel random dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella dan George, 1992).

$$f(t, \theta_{11}, \theta_{12}, \dots, \theta_{1p}, \theta_{21}, \theta_{22}, \dots, \theta_{2p}, \dots, \theta_{Mp})$$

merupakan *joint density* yang karakteristik densitas marginalnya untuk θ_{11} adalah

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

Algoritma *Gibbs Sampling*

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 $\theta_{mk} \sim f(\theta_{mk} | \boldsymbol{\theta}_{-mk}, \boldsymbol{t})$
   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.

Model *mixture* pada regresi merupakan pengembangan bentuk model *mixture* serta tidak bisa dilepaskan dari adanya *mixture* distribusi.

Model *mixture* regresi *survival* dengan fungsi densitasnya tersusun dari distribusi *survival*nya, dengan asumsi bahwa fungsi densitas adalah linier *mixture*, maka bentuk model *survival* dengan *M* komponen adalah

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

Berdasarkan klasifikasi daerah tempat tinggal responden yaitu perdesaan dan perkotaan, maka model *mixture* dengan dua komponen adalah

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

$p(t|\theta_1)$: 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

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.

Ada hubungan 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 (Mc Mahon,1999)

Penduduk Usia Sekolah

Menurut jenjangnya terbagi menjadi lima, sbb: (BPS, 2006)
a. Usia 2-6 tahun, masuk dalam kelompok anak usia dini
b. Usia 7-12 tahun, masuk dalam kelompok usia SD
c. Usia 13-15 tahun, masuk dalam kelompok usia SLTP
d. Usia 16-18 tahun, masuk dalam kelompok usia SLTA
e. Usia 19-24 tahun, masuk dalam kelompok usia sekolah di perguruan tinggi.

Angka Lama Sekolah (ALS)

Lama sekolah didefinisikan sebagai lamanya pendidikan yang ditempuh seseorang dimana tidak memperhatikan cepat atau lambatnya seseorang dalam menempuhnya dari waktu yang telah ditargetkan.

Rumus penghitungan angka lama sekolah sesuai BPS, UNDP (2004), dan Bappenas yaitu

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

Tahun_Konversi : Tahun Pendidikan yang Ditamatkan

− SD	= 6 tahun	- D3	= 15 tahun
− SMP	= 9 tahun	- D4/S1	= 16 tahun
− SMA	= 12 tahun	- S2	= 18 tahun
− D1	= 13 tahun	- S3	= 21 tahun
− D2	= 14 tahun		

Rata-Rata Lama Sekolah (RLS)

Rata-rata lama sekolah atau *Mean Years of Schooling* didefinisikan sebagai jumlah tahun yang digunakan penduduk dalam menjalani pendidikan formal. MYS di suatu daerah didefiniskan sebagai jumlah angka lama sekolah dari setiap warga di daerah tersebut dibagi dengan jumlah warga di daerah tersebut.

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

Sumber Data

raw data SUSENAS 2014 Provinsi Papua Barat.

- ► data angka lama sekolah
- ► *failure event* yaitu apabila responden lulus pada jenjang pendidikan setingkat SMP serta berhasil memiliki ijazah pada rentang umur 16-24 tahun.

Variabel Respon

- responden tidak pernah bersekolah
- mengundurkan diri / *drop out*
- responden belum mendapatkan ijazah SMP/sederajat melebihi batas penelitian.

data tersensor (t=0) tidak mengalami *failure event.*

- responden yang dalam batas periode penelitian mendapatkan ijazah SMP/sederajat.

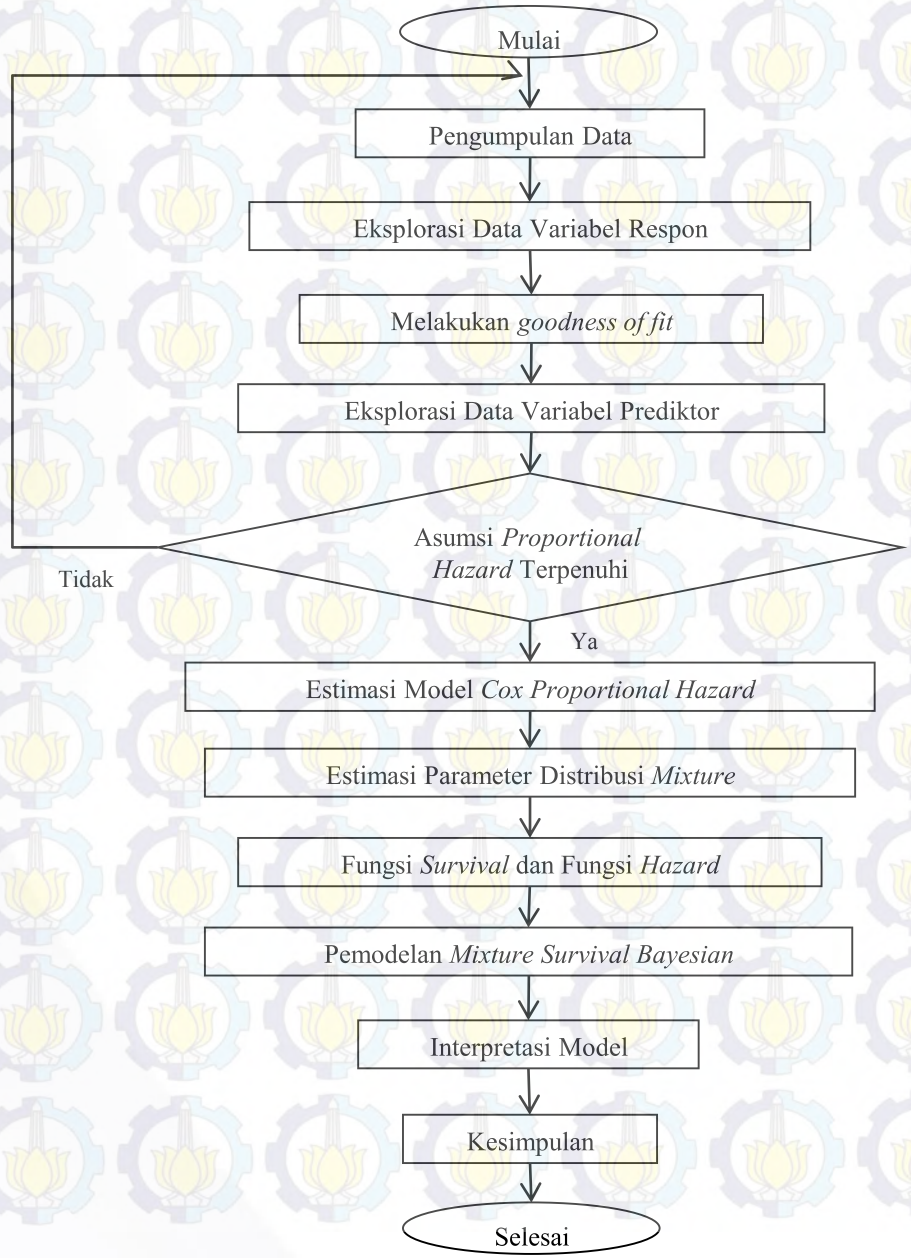
data tidak tersensor (t=1) mengalami *failure event*.

Variabel Prediktor

Jenis Kelamin (X_1)	Status Bekerja (X_2)	Status Perkawinan (X_3)
Dikategorikan: “1” : laki-laki “2” : perempuan	Dikategorikan: “1” : bekerja “2” : tidak bekerja	Dikategorikan: “1” : tidak kawin “2” : kawin (meliputi: kawin, cerai hidup, atau cerai mati)

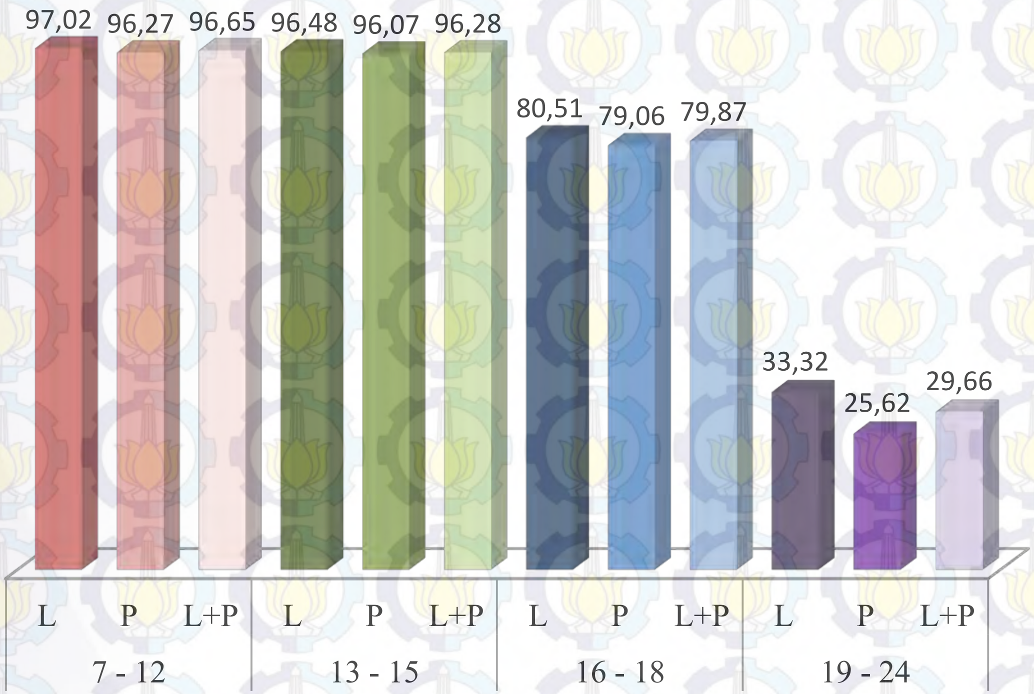
Tingkat Pendidikan KRT (X_4)	Jumlah ART (X_5)	Rata2 Pengeluaran Perkapita sebulan KRT (X_6)
Dikategorikan: “1” : tidak tamat SD “2” : tamat SD “3” : tamat SMP, “4” : tamat SMA ke atas	jumlah anggota rumah tangga yang berada dalam satu rumah tangga (orang)	rata-rata pengeluaran baik makanan maupun non makanan dalam sebulan yang dikeluarkan oleh rumah tangga per kapita (juta rupiah)

- ✓ *pre-processing* data
- ✓ Identifikasi variabel yang mempunyai pengaruh signifikan terhadap lama sekolah di Provinsi Papua Barat dan klasifikasi wilayah dengan regresi *cox proportional hazard*
  - Eksplorasi data variabel lama sekolah menurut perdesaan dan perkotaan
  - Melakukan *goodness of fit* masing-masing distribusi data lama sekolah menurut pedesaan dan perkotaan
  - Eksplorasi data variabel prediktor (*X*) menurut pedesaan dan perkotaan di Provinsi Papua Barat
  - Estimasi parameter model regresi *cox proportional hazard* (uji *proportional hazard* pada variabel prediktor, uji serentak untuk parameter model regresi *cox proportional hazard,* uji parsial untuk parameter model regresi *cox proportional hazard*)
- ✓ Estimasi parameter distribusi mixture berdasarkan distribusi data pada hasil *goodness of fit* dengan MCMC teknik *Gibbs Sampler*
- ✓ Nilai parameter hasil estimasi parameter regresi *cox proportional hazard* digunakan sebagai dasar pembentukan distribusi *prior* (*pseudo prior*) serta hasil estimasi parameter distribusi mixture untuk menyusun dan mengestimasi *mixture* regresi *survival* pada data lama sekolah yang disusun berdasarkan klasifikasi wilayah.



Mulai
Pengumpulan Data
Eksplorasi Data Variabel Respon
Melakukan *goodness of fit*
Eksplorasi Data Variabel Prediktor
Asumsi *Proportional Hazard* Terpenuhi
Tidak
Ya
Estimasi Model *Cox Proportional Hazard*
Estimasi Parameter Distribusi *Mixture*
Fungsi *Survival* dan Fungsi *Hazard*
Pemodelan *Mixture Survival Bayesian*
Interpretasi Model
Kesimpulan
Selesai

Angka Partisipasi Sekolah



Pada kelompok usia 16-18 & 19-24 tahun keduanya masih didominasi oleh penduduk laki-laki yang lebih besar persentase partisipasi sekolahnya.

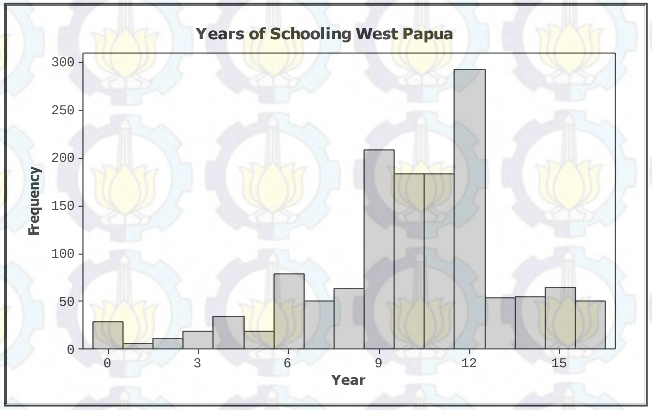
Angka Partisipasi Sekolah

Perkotaan

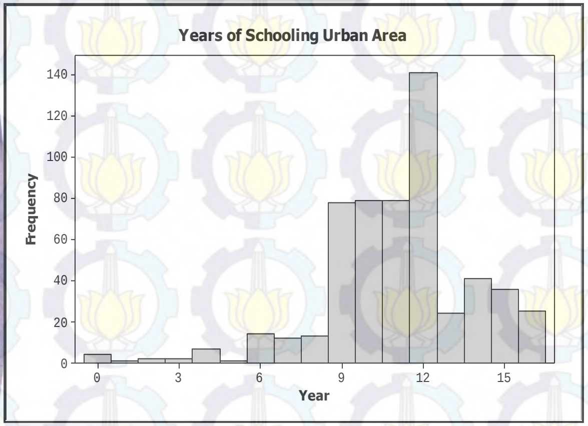
Variabel	Minimum	Maksimum	Mean	Std. Deviasi
Lama Sekolah Penduduk Usia 16 - 24 tahun	0	16	11,08	2,71
Jumlah ART	2	18	5,89	2,49
Rata-rata Pengeluaran Perkapita Sebulan (dlm jutaan rupiah)	0,2594	5,9998	1,0637	0,7133

Perdesaan

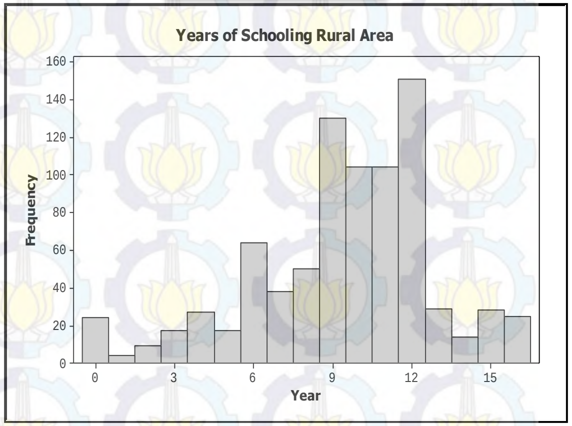
Variabel	Minimum	Maksimum	Mean	Std. Deviasi
Lama Sekolah Penduduk Usia 16 - 24 tahun	0	16	9,43	3,42
Jumlah ART	2	20	5,92	2,47
Rata-rata Pengeluaran Perkapita Sebulan (dlm jutaan rupiah)	0,1448	8,7548	0,5572	0,4638



Perkotaan



Perdesaan

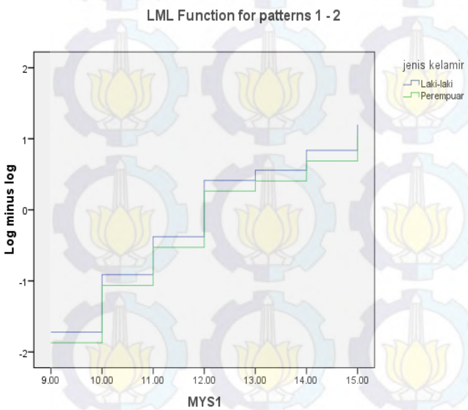


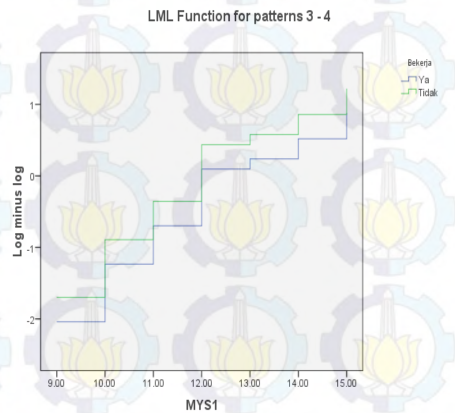
Goodness of fit

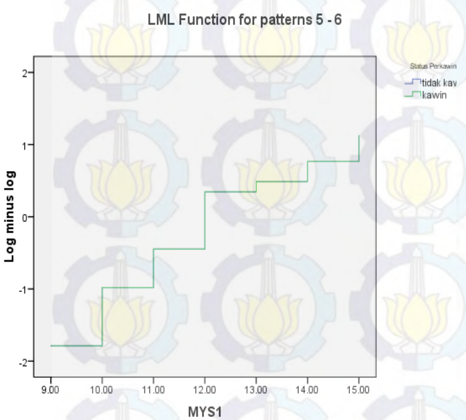
Distribusi	Papua Barat	Perkotaan	Perdesaan	Nilai Kritis
Normal	25,536	9,461	14,012	2,5018
Weibull	74,280	16,566	57,506	2,5018
Lognormal	120,470	29,102	83,742	2,5018
Log-Logistik	97,523	18,360	75,025	2,5018
Gamma	99,292	21,998	71,358	2,5018

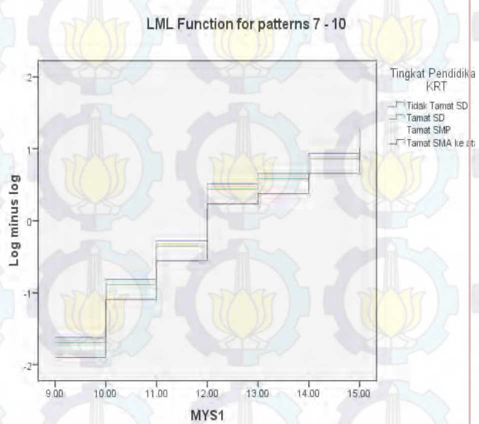
Asumsi Proportional Hazard

Perkotaan

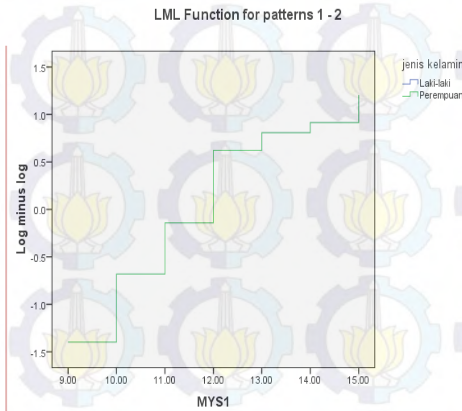


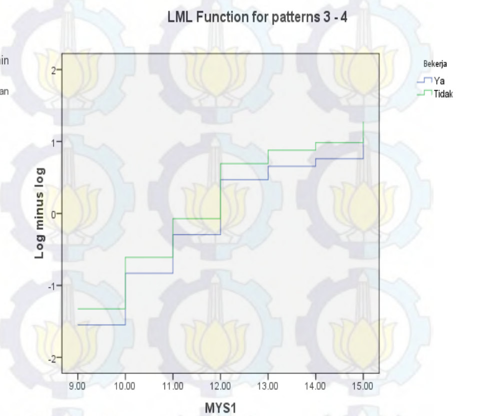


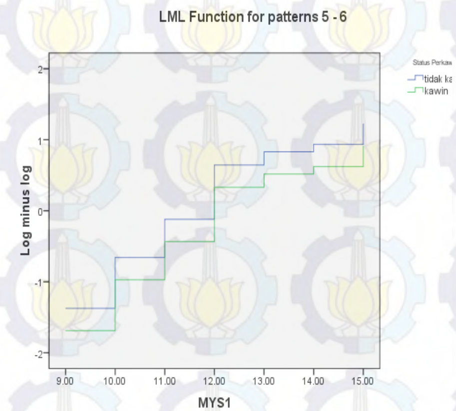


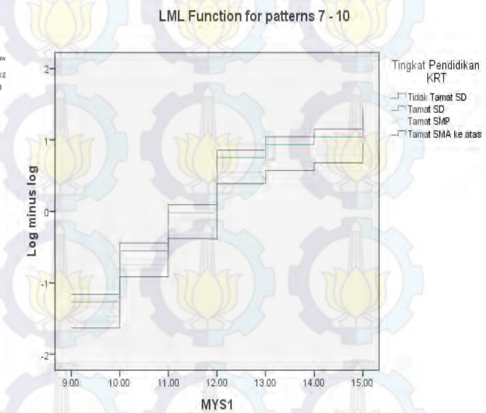


Perdesaan









Plot Fungsi ln(-ln $S(t)$)

Pemilihan Model Terbaik

Perkotaan

Step	Variabel dalam Model	-2Ln 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

Perdesaan

Step	Variabel dalam Model	-2Ln 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

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

Fungsi *Survival* Distribusi *Mixture* 2 Subpopulasi $S(t)=\pi\,S_1(t)+(1-\pi)\,S_2(t),$

Model *proportional hazards* untuk *mixture survival* $h(t)=\pi\,h_1(t)+(1-\pi)\,h_2(t),$

Fungsi Densitas Distribusi *Mixture Weibull*

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

$\lambda > 0$ dan $\gamma > 0$

$0 < t \leq \infty$

Fungsi *Survival* Model *Mixture Weibull*

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

Fungsi *Hazard* Model *Mixture Weibull*

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

Model *Mixture Weibull Proportional Hazard*

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

Distribusi Prior dan Joint Posterior

Distribusi *prior* : *prior conjugate*, pseudo *prior*, dan *informative prior*

$$p(\lambda,\gamma,\pi \mid t) \propto l(t \mid \lambda,\gamma,\pi)\times p(\lambda)\times p(\gamma)\times p(\pi)$$

dimana

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

$p(\gamma) = \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}\gamma^{\alpha-1}\exp(-\dfrac{\gamma}{\beta})$ Gamma(α,β), $\gamma > 0$

$p(\lambda) = \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\dfrac{\lambda}{\beta})$ Gamma(α,β), $\lambda > 0$

$p(\pi) = \dfrac{\Gamma\left(\sum_m \alpha_m\right)}{\prod_m \Gamma(\alpha_m)}\prod_m \pi_m^{\alpha_m-1}$ Dirichlet (α[ ]), $0 < \pi_{\mathrm{m}} < 1$ $\sum_m \pi_m = 1$

Estimasi Parameter Model *Mixture Survival*

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

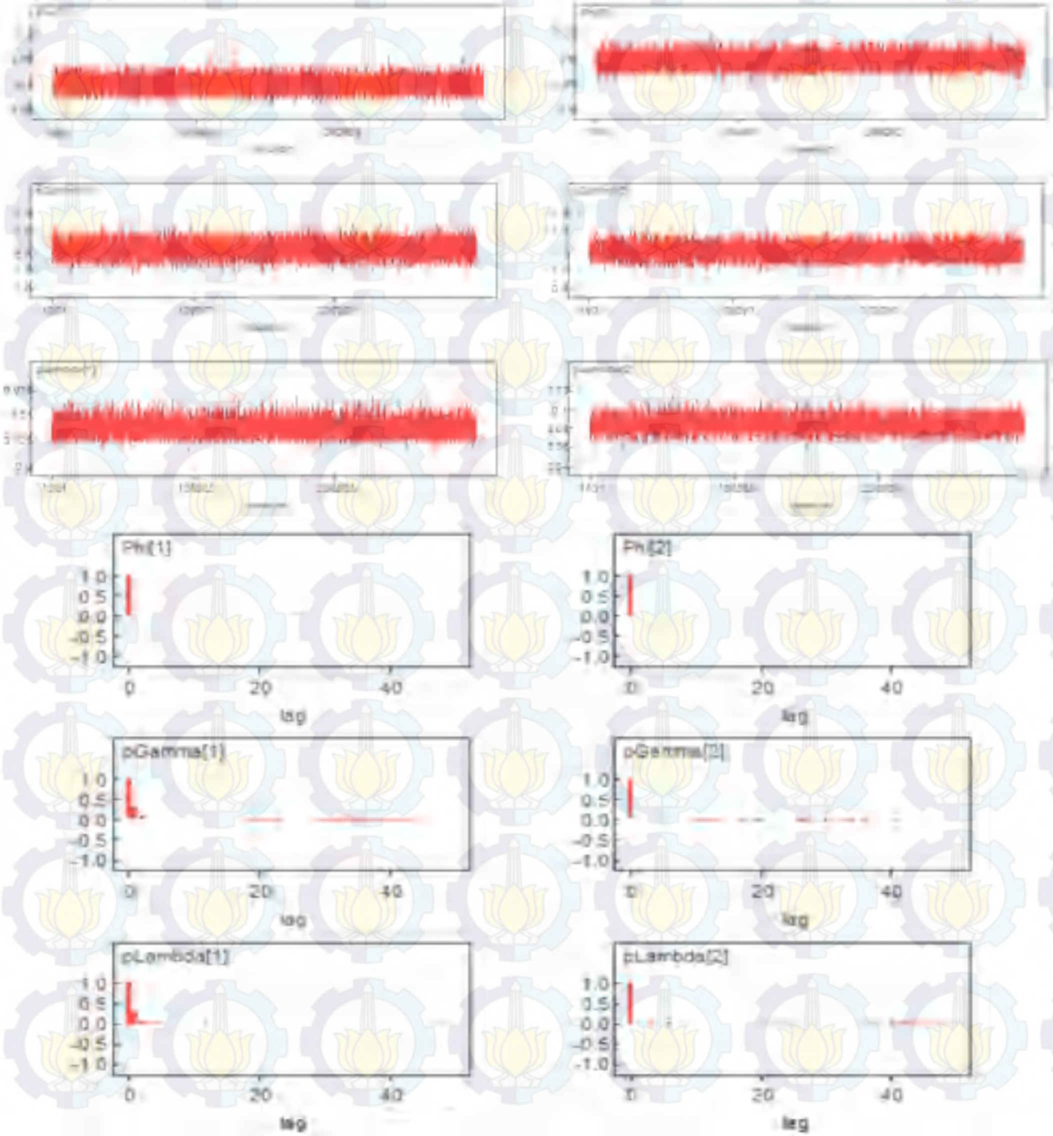
Densitas Distribusi *Mixture Weibull* data lama sekolah

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

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

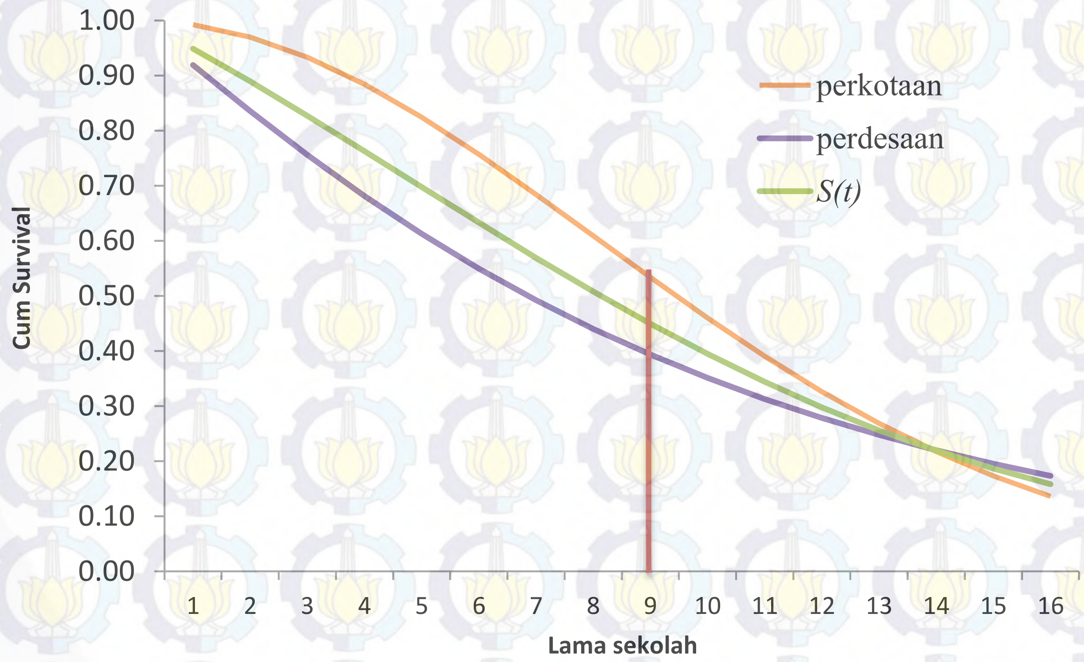
Iterasi 300.000, burn in 1000, thin 25 sehingga jumlah sampel sebanyak 11.960

Monitoring Konvergensi Algoritma



Fungsi Survival Model *Mixture Weibull*

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



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

Estimasi Parameter Model *Mixture Weibull Proportional Hazard*

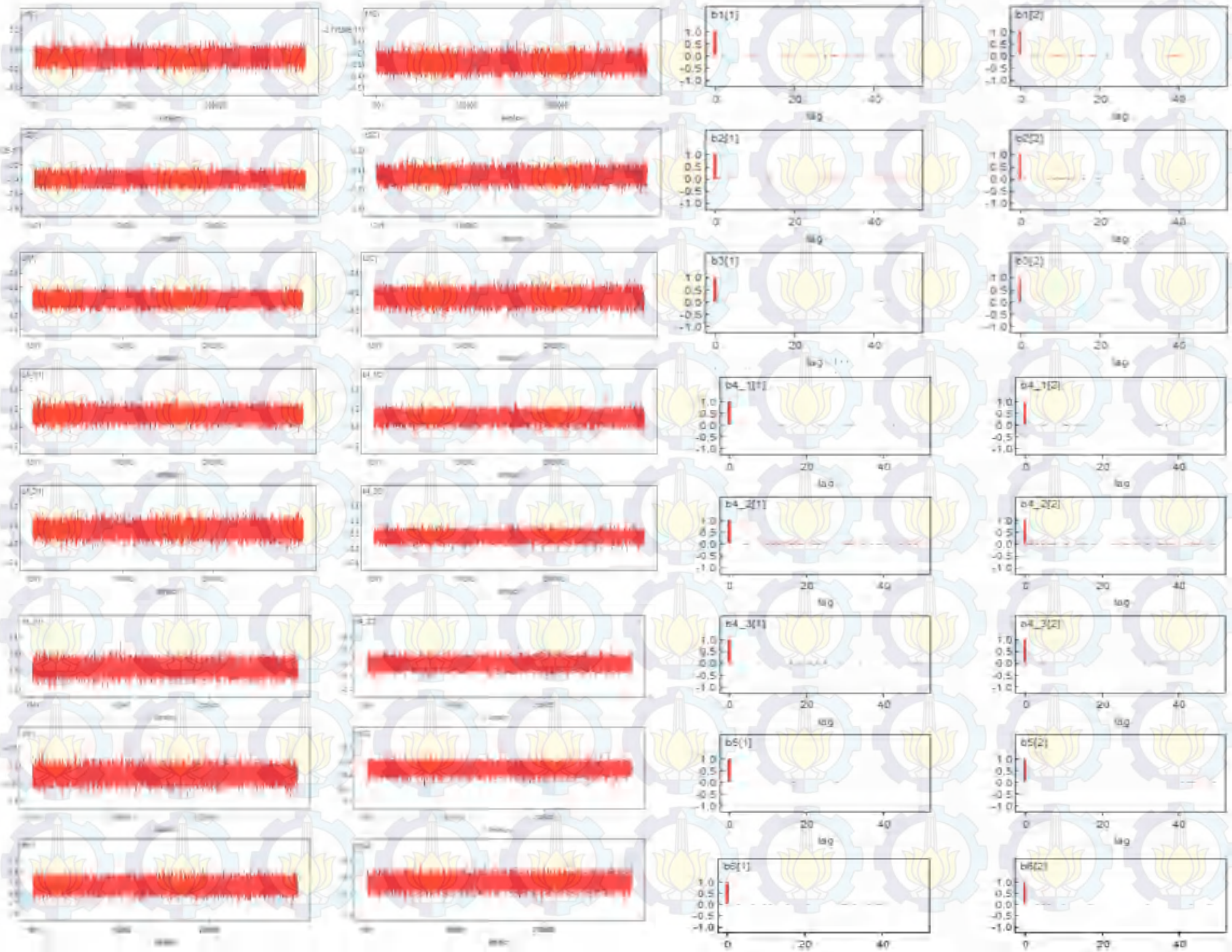
Parameter	Mean	exp (B)	Std.Dev	2,5%	Median	97,5%
b1[1]	-0,0832	0,9202	0,0570	-0,1949	-0,0834	0,0272
b1[2]	-0,2816	0,7546	0,0535	-0,3860	-0,2812	-0,1764
b2[1]	-0,3921	0,6756	0,0624	-0,5158	-0,3918	-0,2701
b2[2]	-0,4617	0,6302	0,0568	-0,5722	-0,4618	-0,3496
b3[1]	-0,9758	0,3769	0,0693	-1,1110	-0,9759	-0,8386
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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
b4_2[2]	-0,0287					
b4_3[1]	0,0385					
b4_3[2]	-0,0124					
b5[1]	-0,3251					
b5[2]	-0,3783					
b6[1]	-0,5411					
b6[2]	-0,6546					

Perkotaan:
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).

Perdesaan:
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)

Iterasi 300.000, burn in 1000, thin 25 sehingga jumlah sampel sebanyak 11.960

Monitoring Konvergensi Algoritma



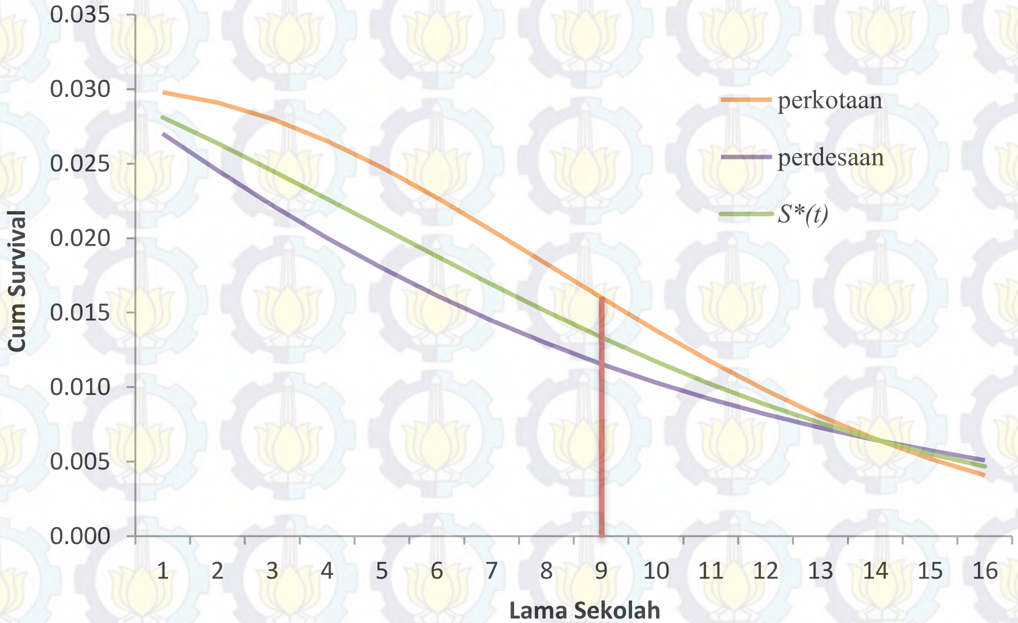
Model *Mixture Weibull Proportional Hazard*

Daerah Perkotaan

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

Daerah Perdesaan

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



Intepretasi Model

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 bersekolah di daerah perkotaan dibandingkan di perdesaan masih konsisten hingga pada level SMA

Perkotaan

Variabel status bekerja (X_2) menunukkan bahwa nilai $\hat{\beta}_2$ = -0,3921 sehingga $\exp(\hat{\beta}_2)$ = 0,6756. Resiko penduduk yang bekerja lebih lama sekolahnya adalah sebesar 0,6756 kalinya penduduk yang tidak bekerja atau 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

Perdesaan

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 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.

KESIMPULAN DAN SARAN

Kesimpulan

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.

Saran

Sebagai bahan pertimbangan pada Pemerintah Daerah Provinsi Papua Barat, di perdesaan perlu mendapatkan perhatian yang serius misalnya dengan menambah fasilitas pendidikan, peningkatan mutu dan kuantitas tenaga pengajar beserta penghargaan yang layak terhadap tenaga pengajar.

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. (2015). Indeks *Pembangunan Manusia Metode Baru*. BPS: Jakarta

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

Brunello,G. dan Checchib, D. (2005), "School quality and family background in Italy", *Economics of Education Review,* Vol. 24, Hal. 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", *of 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,* B34, 187-220.

Cox, D., & 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.

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.

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.

DAFTAR PUSTAKA

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

Iriawan, N. (2002), Studi Tentang Model *Mixture*s 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.

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

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.

DAFTAR PUSTAKA

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.

Terima Kasih