



TESIS - SS142501

**SMALL AREA ESTIMATION PADA KASUS RESPON
MULTINOMIAL DENGAN PENDEKATAN
HIERARCHICAL BAYES
(Studi Kasus : Proporsi Lama Sekolah
di Kabupaten Keerom)**

FIELDRI SINAGA
NRP. 06211650010021

DOSEN PEMBIMBING
Dr. Dra. Agnes Tuti Rumiati, M.Sc.
Dr. Ir. Setiawan, M.S.

PROGRAM MAGISTER
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018



TESIS - SS142501

SMALL AREA ESTIMATION PADA KASUS RESPON
MULTINOMIAL DENGAN PENDEKATAN ***HIERARCHICAL BAYES***
(Studi Kasus: Proporsi Lama Sekolah di Kabupaten Keerom)

FIELDRI SINAGA

NRP. 06211650010021

DOSEN PEMBIMBING :

Dr. Dra. Agnes Tuti Rumiati, M.Sc

Dr. Ir. Setiawan, M.S

PROGRAM MAGISTER

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2018



THESIS - SS142501

***SMALL AREA ESTIMATION USING HIERARCHICAL BAYES
METHODS ON MULTINOMIAL RESPONSE DATA
(Case in Proportion of Years Schooling in Keerom Regency)***

FIELDRI SINAGA

NRP. 06211650010021

SUPERVISOR :

Dr. Dra. Agnes Tuti Rumiati, M.Sc

Dr. Ir. Setiawan, M.S

PROGRAM OF MAGISTER

DEPARTMENT OF STATISTICS

FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCES

INSTITUTE OF TECHNOLOGY SEPULUH NOPEMBER

SURABAYA

2018

**SMALL AREA ESTIMATION PADA KASUS RESPON MULTINOMIAL
DENGAN PENDEKATAN *HIERARCHICAL BAYES*
(Studi Kasus : Proporsi Lama Sekolah di Kabupaten Keerom)**

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

oleh :
FIELDRI SINAGA
NRP. 06211650010021

Tanggal Ujian : 26 Juli 2018
Periode Wisuda : September 2018

Disetujui oleh:

1. Dr. Dra. Agnes Tuti Rumiati, M.Sc.
NIP: 19570724 198503 2 002

(Pembimbing I)

2. Dr. Ir. Setiawan, M.S.
NIP: 19601030 198701 1 001

(Pembimbing II)

3. Dr. Vita Ratnasari, S.Si., M.Si.
NIP: 19709010 199702 2 001

(Penguji)

4. Dr. Drs. Agus Suharsono, M.S.
NIP: 19580823 198403 1 003

(Penguji)

Dekan

Fakultas Matematika, Komputasi dan Sains Data
Institut Teknologi Sepuluh Nopember

Prof. Dr. Basuki Widodo, M.Sc.
NIP: 19650605 198903 1 002

*"Sekalipun aku berjalan dalam lembah kekelaman, aku tidak takut bahaya,
sebab Engkau besertaku; gada-MU dan tongkat-Mu, itulah yang
menghibur aku." (Mazmur 23:4)*

Kupersembahkan kepada:

Bapak Pionel Sinaga dan Ibu Tudaini Sirait

Kedua orang tua ku yang luar biasa

Evalina D. Malau

Istriku tercinta

Graleezanetha F. Sinaga

Gwyvanetha F. Sinaga

Putri-putriku tersayang

Terimakasih atas segala doa, cinta, dukungan dan semangatnya.

(Halaman ini sengaja dikosongkan)

***SMALL AREA ESTIMATION PADA KASUS RESPON MULTINOMIAL
DENGAN PENDEKATAN HIERARCHICAL BAYES
(Studi Kasus: Proporsi Lama Sekolah di Kabupaten Keerom)***

Nama Mahasiswa : Fieldri Sinaga
NRP : 06211650010021
Dosen Pembimbing : Dr. Dra. Agnes Tuti Rumiati, M.Sc
Co-Pembimbing : Dr. Ir. Setiawan, M.S

ABSTRAK

Penelitian ini bertujuan untuk mengestimasi proporsi lama sekolah tingkat kecamatan di Kabupaten Keerom menggunakan pendekatan *Hierarchical Bayes* berdasarkan model *Multinomial Logit*, dan membandingkannya dengan estimasi langsung. Otonomi daerah mendorong pemerintah daerah untuk memiliki informasi yang dapat diandalkan dan detail di daerah kecil. Penghitungan lama sekolah level kecamatan tidak dapat dilakukan secara langsung karena kurang sampel. Salah satu cara untuk mengatasi hal tersebut adalah melalui pendekatan estimasi tidak langsung *Small Area Estimation* (SAE). Salah satu model SAE yang dapat digunakan adalah model *Hierarchical Bayes* (HB). Model *Multinomial Logit* digunakan sebagai model penghubung. Model HB memiliki kelebihan dibandingkan dengan model SAE lainnya karena menghasilkan MSE lebih rendah, varian posterior terkecil, dan dapat menangani variabel respon diskrit. Estimasi proporsi lama sekolah menggunakan model HB menunjukkan beberapa karakteristik, yaitu konvergen distribusi posterior, koefisien variasi lebih rendah dari estimasi langsung.

Kata Kunci: *Hierarchical Bayesian, Small Area Estimation, Gibbs Sampling, Respon Multinomial*

(Halaman ini sengaja dikosongkan)

**SMALL AREA ESTIMATION USING HIERARCHICAL BAYES METHODS
ON MULTINOMIAL RESPONSE DATA**

(Case in proportion of Years Schooling (YS) in Keerom Regency)

By : Fieldri Sinaga
Student Identify Number : 06211650010021
Supervisor : Dr. Dra. Agnes Tuti Rumiati, M.Sc
Co-Supervisor : Dr. Ir. Setiawan, M.S

ABSTRAK

The purpose of this study is to estimate the proportion of years schooling at sub-district level in Keerom District using Hierarchical Bayes approach based on Multinomial Logit model and compare it with direct estimation. Regional autonomy encourages local governments to have reliable and detailed information in small areas. Estimation of years schooling for sub-district level can't be done directly because of lack of sample. One of the way to overcome this is through the indirect estimation approach of Small Area Estimation (SAE). One of the SAE models that can be used is the Hierarchical Bayes (HB) model. The Multinomial Logit model is used as the connecting model. The HB model has advantages over other SAE models, as it produces a smaller MSE, produces the smallest posterior variance, and can handle discrete response variables. The proportion of years schooling estimation using HB model shows some characteristic, that is convergent posterior distribution and the coefficient of variance is smaller than the direct estimate.

Keywords: Hierarchical Bayesian, Small Area Estimation, Gibbs Sampling, Multinomial Response

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah pencipta langit dan bumi, karena hanya karena kasih dan karunia-Nya maka tesis dengan judul *Small Area Estimation* pada Kasus Respon Multinomial dengan Pendekatan *Hierarchical Bayes* (Studi kasus: Proporsi Lama Sekolah di Kabupaten Keerom) dapat diselesaikan. Tesis ini merupakan hasil penelitian yang diajukan sebagai salah satu syarat untuk memperoleh gelar Magister Sains (M.Si) pada Program Studi Magister Jurusan Statistika Fakultas Matematika, Komputasi dan Sains Data Institut Teknologi Sepuluh Nopember Surabaya.

Pada kesempatan ini, penulis mengucapkan terimakasih dan penghargaan yang sebesar-besarnya kepada:

1. Badan Pusat Statistik (BPS) yang telah memberikan kesempatan dan Lembaga Pengelolaan Dana Pendidikan (LPDP) yang telah memberikan beasiswa untuk melanjutkan program studi S2 di ITS.
2. Ibu Dr. Dra. Agnes Tuti Rumiati, M.Sc. dan Bapak Dr. Ir. Setiawan, M.S. sebagai dosen pembimbing atas segala bimbingan, dan saran, serta motivasi yang diberikan selama penyusunan tesis ini.
3. Ibu Dr. Vita Ratnasari, S.Si., M.Si. dan Bapak Dr. Dra. Agus Suharsono, M.S. selaku dosen penguji yang banyak memberikan saran dan koreksi atas penulisan tesis ini.
4. Bapak Ketua Jurusan Statistika, Bapak Ketua Program Studi Pascasarjana Statistika ITS beserta jajarannya atas fasilitas yang disediakan dan arahan selama proses studi.
5. Bapak Dr.rer.pol. Heri Kuswanto, S.Si., M.Si. selaku dosen wali, dan seluruh Bapak/Ibu dosen pengajar yang telah memberikan ilmu dan pengamalan yang bermanfaat kepada penulis.
6. Teman-teman BPS angkatan 10: Rafael, Umam, Fendi, Suprpto, Soni, Aniq, Ratna, Tika, Reny dan Imra, serta *special thanks for* Ratih, Mike, Parih dan Taufik. Terimakasih untuk segala bantuan, kebersamaan, kelucuan, dan kekompakkan selama studi di ITS.

7. Teman-teman reguler angkatan 2016. Senang bertemu kalian dan terimakasih untuk segala keseruannya.
8. Mas Rindang dan Mba Ayuningtyas. Terimakasih atas bantuan ilmu dan diskusi yang mencerahkan.
9. Semua pihak yang telah membantu penyelesaian tesis ini.

Akhir kata, dengan segala kerendahan hati, penulis menyadari bahwa tesis ini jauh dari sempurna. Segala kritik, saran yang bersifat membangun sangat diharapkan demi perbaikan tesis ini. Penulis berharap semoga tesis ini bermanfaat untuk semua pihak yang memerlukan.

Surabaya, Agustus 2018

Fieldri Sinaga

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK	iii
ABSTRACT	v
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	6
1.3. Tujuan Penelitian	6
1.4. Manfaat Penelitian	6
1.5. Batasan Masalah	6
BAB 2 TINJAUAN PUSTAKA	7
2.1. <i>Small Area Estimation</i> (SAE)	7
2.1.1. Model Berbasis Area (<i>Basic Area Level Model</i>)	9
2.1.2. Model Berbasis Unit (<i>Basic Unit Level Model</i>)	11
2.2. Konsep Bayesien pada SAE	11
2.2.1. Metode HB dalam SAE	13
2.2.2. HB untuk Data Binari	15
2.2.3. Distribusi <i>Prior</i>	16
2.2.4. <i>Marcov Chain Monte Carlo</i> (MCMC)	17
2.2.5. <i>Gibbs Sampling</i>	20
2.3. Distribusi Multinomial	21
2.4. WinBUGS	24
2.5. Koefisien Variasi	24
2.6. Estimasi <i>Mean Square Error</i> (MSE)	25

2.6.	Rata-rata Lama Sekolah (RLS)	26
2.7.	Survei Sosial Ekonomi Nasional (Susenas)	27
2.8.	Penelitian terkait Penentuan Variabel Bebas	29
BAB 3	METODOLOGI PENELITIAN	31
3.1.	Sumber Data	31
3.2.	Metode Pengumpulan Data	31
3.3.	Variabel Penelitian	32
3.4.	Metode untuk Kajian Teori	33
3.5.	Aplikasi Model	34
BAB 4	HASIL DAN PEMBAHASAN	37
4.1.	Estimasi Parameter <i>Multinomial Logit Mixed Model</i> dengan Pendekatan <i>Hierarchical Bayes</i>	37
4.2.	Gambaran Umum Kabupaten Keerom	42
4.3.	Estimasi Proporsi Lama Sekolah menggunakan Metode <i>Hierarchical Bayes</i> (HB) dengan <i>Multinomial Logit Mixed Model</i>	46
4.4.	Perbandingan Koefisien Variasi (CV)	50
4.5.	Perbandingan <i>Mean Square Error</i> (MSE)	51
4.6.	Pembahasan	52
BAB 5	KESIMPULAN DAN SARAN	55
5.1.	Kesimpulan	55
5.2.	Saran	55
	DAFTAR PUSTAKA	57
	LAMPIRAN	61

DAFTAR TABEL

Tabel 4.1. Topografi, Jumlah Sekolah dan Jarak berdasarkan Kecamatan di Kabupaten Keerom Tahun 2015	44
Tabel 4.2. Jumlah Sampel Penduduk Usia 25 tahun ke atas menurut Kategori Lama Sekolah.....	45
Tabel 4.3. Estimasi Langsung Proporsi Lama Sekolah berdasarkan Kecamatan di Kabupaten Keerom, Tahun 2015	45
Tabel 4.4. Estimasi Proporsi Lama Sekolah menurut Kecamatan menggunakan Metode HB, Tahun 2015	49

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

Gambar 2.1.	Diagram Alur Algoritma <i>Gibbs Sampling</i>	21
Gambar 4.1.	Rata-rata Lama Sekolah Kabupaten Keerom Tahun 2012 – 2015	42
Gambar 4.2.	Jumlah Sampel dan Persentase Sampel Penduduk Usia 25 tahun ke atas terhadap Sampel Susenas	43
Gambar 4.3.	<i>Plot</i> distribusi <i>Posterior</i> Parameter p_1^{HB} , p_2^{HB} , dan p_5^{HB}	47
Gambar 4.4.	Taksiran Parameter β_{03} dan β_{15} dalam estimasi proporsi lama sekolah	48
Gambar 4.5.	Perbandingan CV Proporsi Lama Sekolah antara Estimasi Langsung dan Metode HB pada a) Kategori 1, b) Kategori 2, c) Kategori 3, d) Kategori 4, e) Kategori 5, dan f) Kategori 6	51
Gambar 4.6.	Gambar 4.6. Perbandingan MSE Proporsi Lama Sekolah antara Estimasi Langsung dan Metode HB pada a) Kategori 1, b) Kategori 2, c) Kategori 3, d) Kategori 4, e) Kategori 5, dan f) Kategori 6	52

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1.	Syntax Metode <i>Hierarchical Bayes</i>	61
Lampiran 2.	Hasil Estimasi Parameter menggunakan Metode HB	63

(Halaman ini sengaja dikosongkan)

BAB 1.

PENDAHULUAN

1.1. Latar Belakang

Pendidikan merupakan salah satu program pembangunan nasional yang berperan penting dalam meningkatkan kualitas hidup dan mencapai kemajuan bangsa. Oleh karena itu, pemerintah terus mengupayakan berbagai kebijakan untuk memenuhi hak setiap warga negara dalam mendapatkan layanan pendidikan. Pergeseran dari sentralisasi menuju desentralisasi seperti yang diamanatkan UU Nomor 32 Tahun 2004 Tentang Pemerintahan Daerah, memberikan keleluasaan kepada pemerintah daerah menerapkan kebijakan otonomi daerah dengan memberikan hak, wewenang dan kewajiban untuk mengatur urusan pemerintahan dan kepentingan masyarakat setempat sesuai dengan peraturan perundang-undangan. Dengan demikian, pemerintah daerah akan sangat membutuhkan informasi mengenai wilayahnya, baik informasi di tingkat kabupaten/kota maupun domain yang lebih kecil seperti kecamatan.

Pemerintah daerah harus serius melaksanakan 1) UU No. 21 Tahun 2001 Tentang Otonomi khusus Pasal 56 ayat 1, yaitu tanggung jawab Pemerintah Provinsi dalam menyelenggarakan pendidikan pada semua jenjang, jalur dan jenis pendidikan di Propinsi Papua”, dan 2) Peraturan Daerah Nomor 13 Tahun 2016 tentang pembagian dana otonomi khusus dimana alokasi anggaran untuk sektor pendidikan ditetapkan 30 persen dari dana otonomi khusus. Selain itu, tujuan keempat dalam Tujuan Pembangunan Berkelanjutan (TPB) yang harus dicapai adalah menjamin kualitas pendidikan yang adil dan inklusif serta meningkatkan kesempatan belajar seumur hidup untuk semua. Agenda pembangunan nasional dalam Nawacita pada butir kelima juga menegaskan bahwa pemerintah akan memprioritaskan peningkatan kualitas hidup manusia Indonesia melalui dua program, yaitu peningkatan kualitas pendidikan dan pelatihan melalui program “Indonesia Pintar”; dan peningkatan kesejahteraan masyarakat melalui program “Indonesia Kerja” dan “Indonesia Sejahtera”.

Sasaran dari program “Indonesia Pintar” yaitu 1) meningkatkan angka partisipasi pendidikan dasar dan menengah, 2) meningkatkan angka keberlanjutan pendidikan ditandai dengan menurunnya angka putus sekolah dan meningkatnya angka me-

lanjutkan, 3) menurunnya kesenjangan partisipasi pendidikan antar kelompok masyarakat, 4) meningkatnya kesiapan siswa pendidikan menengah untuk memasuki pasar kerja atau melanjutkan ke jenjang pendidikan tinggi, 5) meningkatnya jaminan kualitas pendidikan, tersedianya kurikulum yang andal, dan tersedianya sistem penilaian pendidikan yang komprehensif, dan 6) meningkatnya proporsi siswa SMK yang dapat mengikuti program pemagangan di industri.

Salah satu indikator yang merepresentasikan kondisi suatu wilayah adalah Indeks Pembangunan Manusia (IPM). IPM merupakan salah satu indikator yang dijadikan dasar dalam perencanaan pembangunan. IPM diukur dalam 3 dimensi yaitu dimensi: 1) umur panjang dan sehat, diukur dari angka harapan hidup, 2) pengetahuan, diukur dari harapan lama sekolah dan rata-rata lama sekolah (RLS), serta 3) standar hidup layak, diwakili oleh pengeluaran per kapita (BPS Kabupaten Keerom, 2016). IPM dapat digunakan untuk mengklasifikasikan sebuah wilayah ke dalam kategori maju, berkembang atau terbelakang, serta digunakan untuk mengukur pengaruh dari kebijakan ekonomi terhadap kualitas hidup. Penghitungan IPM dilakukan oleh Badan Pusat Statistik (BPS), yang secara resmi mempublikasikan IPM secara periodik setiap tahun untuk tingkat provinsi dan kabupaten/kota.

Berbagai survei dilakukan dengan tujuan untuk menyediakan data statistik yang lengkap, cepat dan berkesinambungan. Dibandingkan dengan sensus, survei lebih banyak dipilih karena beberapa alasan, diantaranya karena survei membutuhkan biaya yang relatif lebih kecil, membutuhkan waktu dan tenaga yang jauh lebih sedikit, kesalahan dari *sampling error* yang dapat diukur, serta karakteristik atau variabel yang tercakup lebih banyak dan terinci.

Sumber data utama dalam penghitungan IPM menggunakan data Survei Sosial Ekonomi Nasional (Susenas). Estimasi langsung dengan menggunakan data Susenas akan memberikan akurasi yang baik pada level terbatas sesuai desain samplingnya. Susenas dirancang untuk menghasilkan estimasi di tingkat nasional, provinsi dan kabupaten/kota. Akan tetapi, data statistik yang dihasilkan tidak dapat digunakan untuk estimasi pada area yang lebih kecil, seperti kecamatan, dikarenakan jumlah sampel yang tidak memadai. Bila ukuran sampel dalam area kecil berukuran kecil, estimasi langsung tidak mampu memberikan ketelitian yang cukup, sehingga statistik yang

dihasilkan akan memiliki varian yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Prasad dan Rao, 1990).

Estimasi langsung suatu indikator pada area kecil dengan jumlah sampel yang sedikit akan menghasilkan *sampling error* yang besar (Benavent dan Morales, 2016). Estimasi parameter dapat dilakukan sampai unit yang lebih kecil jika dilakukan penambahan sampel dari suatu survei, sehingga dapat memperbaiki reliabilitas dari estimasi secara langsung, tetapi juga mengakibatkan biaya pendataan yang dibutuhkan bertambah, jumlah petugas yang dilatih bertambah, serta waktu yang diperlukan dalam melaksanakan survei akan menjadi lebih lama.

Dengan demikian, diperlukan suatu metode tidak langsung yang mampu mengatasi hal tersebut. *Small Area Estimation* (SAE) merupakan teknik yang sering digunakan, dan diharapkan mampu menghasilkan presisi yang lebih baik dalam mengestimasi suatu parameter pada area kecil, dapat berupa area geografis, seperti provinsi, kabupaten/kota, kecamatan, kelurahan/desa, dan sebagainya, atau dapat juga berupa kelompok sosial demografi atau sering disebut domain seperti jenis kelamin, ras, kelompok umur, dan sebagainya. Estimasi dalam SAE didasarkan pada model, sehingga dibutuhkan informasi tambahan dari variabel yang memiliki hubungan dengan variabel yang sedang diamati yang disebut sebagai variabel penyerta.

Pada tahun 1979, Fay dan Herriot memperkenalkan model persamaan untuk mengestimasi suatu parameter pada wilayah kecil. Hingga saat ini, SAE telah banyak diimplementasikan di berbagai negara dan berbagai bidang, antara lain Rao (2003), Bleuer dkk (2007), Hidioglou (2007) dan Noviani (2016). Model SAE dapat berupa model yang berbasis area maupun unit. Penggunaan model SAE berbasis level unit telah dilakukan antara lain oleh Molina dkk.(2007), Scealy (2010), López-Vizcaíno dkk.(2013), Rumiati (2012) dan Miranti(2015). SAE berbasis level area juga telah banyak dilakukan antara lain oleh Zhou dan You(2008), dan Noviani (2016).

Model SAE adalah model estimasi yang memperhitungkan dua jenis ragam, yaitu 1) ragam peubah respon yang tidak dapat diterangkan seluruhnya oleh hubungan peubah respon dengan informasi tambahan yang disebut model pengaruh tetap, dan 2) ragam spesifik area kecil yang tidak dapat diterangkan oleh informasi tambahan merupakan pengaruh acak area kecil. Oleh karena itu model SAE mengandung dua komponen eror yaitu eror karena model dan eror karena pendugaan

parameter secara langsung. Rataan atau total area kecil dapat dinyatakan sebagai kombinasi linier dari pengaruh tetap dan pengaruh acak (Rumiati, 2012).

Secara umum terdapat beberapa metode SAE yang sering digunakan antara lain *Empirical Best Linier Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB) *Estimation* (Ghosh dan Rao, 1994). Metode EBLUP dengan pendekatan *Linear Mixed Model* (LMM) sering digunakan dalam SAE untuk variabel respon yang bersifat kontinu dan telah diketahui mempunyai efisiensi yang baik (Chandra, Chambers, & Salvati, 2009). Jika variabel respon yang diteliti berasal dari data kategorik (biner atau cacahan), maka penggunaan LMM menjadi kurang tepat (Rao, 2003). EB dan HB merupakan metode yang lebih umum digunakan dalam menangani SAE dengan data kategorik karena metode ini dapat diterapkan pada *Generalized Linear Mixed Model* (GLMM).

SAE dengan metode HB cenderung lebih disukai dibanding pendekatan lainnya karena pembentukan model estimasinya menggabungkan informasi bukan hanya dari data sampel melainkan melibatkan informasi lain seperti data terdahulu, bahkan pengetahuan peneliti sendiri. Selain itu, SAE dengan metode HB mempunyai beberapa keuntungan, antara lain yaitu 1) spesifikasinya yang langsung dan dapat digunakan untuk memodelkan berbagai sumber variasi yang berbeda dan 2) komputasinya relatif lebih mudah dengan adanya metode *Markov Chain Monte Carlo* (MCMC) (Hajarisman, 2013). SAE dengan metode HB misalnya digunakan oleh Noviani (2016), dan Ayuningtyas (2017).

Umumnya, pengembangan model pada SAE dengan pendekatan *Bayes* difokuskan pada data kontinu, padahal data survei seringkali juga berbentuk data cacahan (diskrit) atau kategorik. Model SAE untuk data cacahan kebanyakan mengasumsikan bahwa variabel respon berdistribusi Poisson (Trevisani dan Torelli, 2007). Model untuk data kategorik yang bersifat dikotomus umumnya menggunakan model Binomial dengan distribusi *prior* Beta atau Logit Normal, misalnya Maiti (1997) dan Liu (2009). Selanjutnya model yang dapat diterapkan pada data kategorik yang bersifat politomus yaitu Multinomial.

Molina, Saei dan Lombardia (2007), Scealy (2010), López-Vizcaíno, Lombardia dan Morales (2013), López-Vizcaíno dkk (2015), Rumiati (2012) dan Miranti (2015) mengembangkan model multinomial untuk SAE (Ayuningtyas,

2017). Rumiati (2012) menggunakan pendekatan klasik dengan *Multinomial Logit Mixed Model* dalam estimasi parameter pada respon multinomial, dengan menggunakan *Multinomial Weighted Logit Mixed Model*, dimana untuk mengestimasi parameter digunakan pendekatan *Empirical Bayes*. Miranti (2015) juga menggunakan model SAE berbasis unit dan *Multinomial Logit Mixed Model* sebagai model *linking* yang digunakan untuk estimasi parameter pada pendekatan HB.

Dalam mengestimasi tingkat penyakit asma pada 20 wilayah kesehatan di Provinsi British Columbia, Canada, Zhou dan You (2008) membandingkan koefisien variasi (CV) dari hasil estimasi menggunakan HB. Penelitian Zhou dan You menghasilkan kesimpulan bahwa estimasi menggunakan metode HB berhasil menurunkan nilai CV secara signifikan dari estimasi secara langsung. Ayuningtyas (2017) dalam penelitiannya juga menyimpulkan bahwa estimasi menggunakan model HB memiliki penurunan koefisien variasi secara signifikan dari hasil estimasi langsung pada estimasi proporsi pengangguran di Pulau Kalimantan.

Teknik yang digunakan dalam SAE adalah untuk memperoleh estimasi dengan tingkat presisi yang tinggi pada area kecil dapat digambarkan oleh *Mean Square Error* (MSE) (Baillo dan Molina, 2009). Rao (2007) menjelaskan bahwa ketepatan model dalam SAE dengan kriteria bias sangat sulit dilakukan karena nilai parameter populasi pada wilayah kecil tidak diketahui, sehingga parameter populasi juga diestimasi melalui estimasi dari sampel yang tersedia. Prasad dan Rao (1990), Datta dan Lahiri (2000) serta Datta, Rao, dan Smith (2005) dalam Rao (2007) menggunakan *Taylor linearization* untuk mengestimasi MSE yaitu untuk mengoreksi ketidakpastian akibat menduga parameter populasi menggunakan data sampel. Namun, Rao juga menyebutkan bahwa menduga MSE menggunakan *Taylor linearization* merupakan sesuatu yang kompleks dan sulit. Rao menyebutkan bahwa Jiang, Lahiri, dan Wan (2002) mengusulkan metode *Jackknife* untuk mengoreksi estimasi MSE. Estimasi MSE untuk SAE menggunakan metode *Jackknife* juga digunakan oleh Rumiati (2012), Miranti (2015) dan Ayuningtyas (2017).

Penggunaan model SAE berbasis area dan *Multinomial Logit Mixed Model* dengan pendekatan HB akan diaplikasikan untuk mengestimasi RLS berdasarkan kecamatan di Kabupaten Keerom Provinsi Papua.

1.2. Rumusan Permasalahan

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana memperoleh estimasi parameter dari model SAE berbasis area dan *Multinomial Logit Mixed Model* dengan pendekatan HB?
2. Berapa proporsi lama sekolah pada masing-masing kecamatan di Kabupaten Keerom berdasarkan estimasi langsung dan model SAE?

1.3. Tujuan Penelitian

1. Memperoleh estimasi langsung dan estimasi parameter *Multinomial Logit Mixed Model* dengan pendekatan HB.
2. Mengaplikasikan *Multinomial Logit Mixed Model* dengan pendekatan HB untuk mengestimasi proporsi lama sekolah pada kategori lama sekolah pada tingkat kecamatan di Kabupaten Keerom.
3. Membandingkan estimasi langsung dan estimasi menggunakan *Multinomial Logit Mixed Model* dengan pendekatan HB.

1.4. Manfaat Penelitian

1. Memberikan informasi mengenai metode *Multinomial Logit Mixed Model* dengan pendekatan HB untuk estimasi parameter pada respon multinomial dan berupa proporsi pada wilayah/domain yang lebih kecil.
2. Tersedianya data lama sekolah pada tingkat kecamatan yang dapat dijadikan dasar dalam penanganan dan evaluasi program pembangunan tepat sasaran di bidang pendidikan.

1.5. Batasan Masalah

1. Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari berbagai publikasi yang diterbitkan oleh BPS Kabupaten Keerom.
2. Model SAE yang dibentuk adalah model berbasis level area.

Model estimasi parameter menggunakan iterasi numerik karena penyelesaian persamaan secara analitik model Bayes berbasis data biner sulit ditemukan.

BAB 2

TINJAUAN PUSTAKA

2.1. *Small Area Estimation (SAE)*

Kebutuhan akan data pada area terkecil, misalnya kecamatan, semakin meningkat. Hal ini berkaitan erat dengan kepentingan pembangunan nasional yang mewajibkan pemerintah daerah menyediakan informasi yang lengkap. Dengan melaksanakan kegiatan sensus, maka informasi hingga level area terkecil dapat diperoleh. Sensus seringkali dilaksanakan pada periodisasi yang lama karena membutuhkan sumber daya yang besar. Di Indonesia, sensus dilaksanakan dalam kurun waktu sepuluh tahunan, dan ditujukan terutama untuk mendapatkan informasi dasar dan berfungsi dalam pembentukan kerangka sampel keperluan survei. Untuk tujuan efektifitas, metode survei lebih sering dilakukan sebagai cara mendapatkan suatu penduga parameter.

Pada umumnya, survei yang dilaksanakan oleh Badan Pusat Statistik (BPS) digunakan dan dirancang untuk menduga parameter dari suatu data pada level area tingkat nasional, provinsi maupun kabupaten/kota. Pada area tersebut, sampel yang diperoleh dapat langsung digunakan untuk estimasi karena sesuai dengan desain samplingnya. Estimasi parameter pada area kecil yang didasarkan pada penerapan model rancangan penarikan sampel disebut sebagai penduga langsung, artinya melakukan estimasi parameter populasi suatu area hanya berdasarkan data yang diperoleh dari area tersebut. Nilai pendugaan langsung pada area kecil merupakan penduga tak bias, akan tetapi tidak mampu memberikan ketelitian yang cukup dan memiliki ragam/varian, atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Ha, 2013). Untuk ketersediaan informasi pada level yang lebih kecil dibanding desain samplingnya, maka dikembangkan suatu teknik yang dikenal dengan istilah *Small Area Estimation (SAE)*.

Menurut Setiawan (2004) dalam Wulansari (2015), secara umum kebutuhan penerapan SAE didasari oleh beberapa hal, antara lain:

1. Meningkatnya tuntutan dunia akan informasi statistik pada area yang lebih kecil.
2. Derasnya arus informasi yang sudah menembus batas wilayah.
3. Ketersediaan dana.

Istilah *small area* biasa digunakan untuk menunjukkan suatu bagian dari populasi atau bagian dari suatu wilayah geografis seperti provinsi, kabupaten/kota, kecamatan, kelurahan/desa atau blok sensus. Istilah ini juga dapat digunakan untuk menggambarkan pengelompokan demografis seperti kelompok umur tertentu dari sekelompok penduduk di wilayah geografis yang besar. Suatu wilayah dikategorikan sebagai “*small area*” apabila ukuran sampel pada wilayah tersebut tidak cukup untuk mendukung estimasi langsung dengan ketelitian yang baik (Rao, 2003). Istilah lain yang sering digunakan untuk merujuk “*small area*” antara lain *local area*, *sub-domain*, *small group*, dan *minor domain*.

Metode SAE merupakan metode estimasi tidak langsung (*indirect estimation*) yang bersifat “meminjam” kekuatan (*borrowing strength*) dari pengamatan sampel area yang berdekatan dengan memanfaatkan informasi tambahan yang bertujuan untuk meningkatkan efisiensi (Rao, 2003). Penggunaan model SAE memberikan beberapa keuntungan (Rao, 2003):

1. Diagnostik model dapat digunakan untuk mendeteksi kecocokan dengan data.
2. Pengukuran presisi spesifik area dapat diasosiasikan dengan setiap estimasi pada setiap area kecil.
3. Model linear campuran dengan pengaruh acak area–spesifik tetap dapat dilakukan, demikian juga untuk struktur data yang cukup kompleks misalkan struktur data *time series* atau spasial.
4. Pengembangan metode untuk model pengaruh acak dapat dimanfaatkan untuk mencapai akurasi dalam area kecil.

Terdapat dua asumsi yang mendasari adanya pendugaan area kecil. Pertama, asumsi bahwa keragaman di dalam area kecil variabel respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan yang disebut model pengaruh tetap (*fixed effect model*). Kedua, asumsi keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan yang disebut pengaruh acak area (*random effect*). Gabungan antara kedua model tersebut membentuk model campuran. SAE memiliki dua jenis model dasar yaitu model berbasis level area (*basic area level model*) dan berbasis level unit (*basic unit level model*) (Rao, 2003).

2.1.1. Model Berbasis Level Area (*Basic Area Level Model*)

Model berbasis level area merupakan model yang didasarkan pada ketersediaan data pendukung yang hanya ada untuk level area tertentu. Model SAE berbasis area memuat dua model, yaitu *sampling* dan *linking model*. *Sampling model* dihitung berdasarkan estimasi *sampling error* dari survei langsung. Sementara *linking model* menghubungkan suatu nilai populasi (parameter) dengan variabel tambahan dari spesifik area. Karena estimasi berbasis desain survei dimodelkan secara langsung, maka model level area biasanya menghasilkan estimator yang *design-consistent*.

Walaupun telah bersifat *design-consistent*, model berbasis level area tetap membutuhkan estimasi varians sampling yang tepat yang diestimasi dari desain survei dimana data didapatkan (Liu, 2009). Mengingat varians dari ukuran sampel yang kecil cenderung tidak stabil, mengaplikasikan model level area yang menghendaki estimasi dari varians sampling yang tepat menjadi suatu tantangan tersendiri (Ha, 2013).

Misalkan terdapat suatu populasi dengan ukuran N yang dibagi ke dalam m area kecil, dengan masing-masing area berukuran N_1, N_2, \dots, N_m , yang disebut sebagai area atau domain. Jika ingin diperoleh rata-rata dari setiap area dengan variabel yang diamati pada unit ke- j untuk area ke- i adalah Y_{ij} , maka rata-rata persamaannya dapat dituliskan:

$$\mu_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}, \quad \text{untuk } j=1, 2, \dots, m.$$

Untuk mendapatkan estimasi pada level area tertentu, diambil sampel sebanyak n dari populasi dengan menggunakan metode sampling tertentu. Estimator langsung, misal $\hat{\theta}_i$, merupakan estimator dari μ_i yang menggunakan data sampel. Maka menurut Fay dan Herriot (1979), estimator langsung akan memberikan varians yang besar. Untuk mengurangi varians yang besar tersebut, Fay dan Herriot mengasumsikan bahwa $\theta_i = h(\mu_i)$ untuk beberapa $h(\cdot)$ dapat dihubungkan dengan data pendukung dari suatu spesifik area, misalnya sejumlah p variabel prediktor yang

diukur dari area kecil tersebut, yaitu $\mathbf{z}_i = (z_{1i}, z_{2i}, \dots, z_{pi})^T$. Misalkan $\hat{\theta}_i$ merupakan estimasi langsung dari θ , maka model Fay dan Herriot yang terbentuk adalah:

$$\theta_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i, \text{ untuk } i = 1, 2, \dots, m, \quad (2.1)$$

dengan $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ adalah vektor koefisien regresi berukuran $p \times 1$, p adalah banyaknya variabel prediktor, v_i adalah pengaruh acak area spesifik yang diasumsikan memiliki distribusi $v_i \sim N(0, \sigma_v^2)$.

Untuk membuat inferensi mengenai rata-rata dari variabel yang diamati pada area kecil Y_{ij} dengan menggunakan model (2.1), diasumsikan bahwa penduga langsung dari μ_i tersedia, maka Rao (2003) mengasumsikan bahwa:

$$\hat{\theta}_i = h(\hat{\mu}_i) = \theta_i + e_i, \text{ untuk } i = 1, 2, \dots, m, ! \quad (2.2)$$

dengan e_i adalah *sampling error* yang saling bebas dengan asumsi $e_i \sim N(0, \sigma_\varepsilon^2)$ dengan varians dari sampling, σ_ε^2 , pada model Fay dan Herriot lazimnya diasumsikan diketahui. Dalam aplikasinya, asumsi bahwa komponen varians sampling ini sulit terpenuhi sehingga justru harus diestimasi dari data yang didapatkan.

Dengan mengkombinasikan persamaan (2.1) dan (2.2) diperoleh persamaan linier campuran untuk menduga karakteristik area sebagai berikut:

$$\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i + e_i. \quad (2.3)$$

yang merupakan kasus khusus dari *generalized linear mixed model* dan dikenal sebagai model Fay-Herriot, dimana keragaman peubah respon di dalam area kecil diasumsikan dapat diterangkan oleh hubungan peubah respon dengan informasi tambahan yang disebut sebagai model pengaruh tetap. Selain itu terdapat komponen keragaman spesifik area kecil yang tidak dapat diterangkan oleh informasi tambahan dan disebut sebagai komponen pengaruh acak area kecil. Gabungan dari dua asumsi tersebut membentuk model pengaruh campuran (*mixed model*).

Model (2.3) tidak tepat untuk langsung digunakan jika parameter θ_i bukan merupakan fungsi linier dari total Y_i sehingga dalam hal ini model sampling tidak tepat untuk dipasangkan langsung dengan model penghubung. Menurut Rao (2003), untuk

menangani masalah model sampling yang tidak dapat dipasangkan langsung dengan model penghubung dapat diselesaikan dengan pendekatan HB.

2.1.2. Model Berbasis Level Unit (*Basic Unit Level Model*)

Model SAE berbasis level unit merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon, misal $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ untuk setiap elemen ke- j pada area ke- i . Variabel respon y_{ij} diasumsikan memiliki hubungan dengan x_{ij} melalui model persamaan (2.4):

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, m \text{ dan } j = 1, \dots, N, \quad (2.4)$$

dengan $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$, p adalah banyaknya variabel prediktor, m merupakan banyaknya individu atau anggota rumah tangga di area ke- i , N merupakan banyaknya area, v_i adalah pengaruh acak area dan e_{ij} adalah peubah acak bersifat bebas terhadap v_i , dan keduanya diasumsikan berdistribusi $v_i \sim N(0, \sigma_v^2)$ dan $e_{ij} \sim N(0, \sigma_\varepsilon^2)$.

Perbedaan mendasar pada kedua model tersebut yaitu pada penggunaan data pendukung yang tersedia. Pada model SAE berbasis area, data pendukung yang tersedia hanya untuk level area tertentu. Model ini menghubungkan estimator langsung dengan variabel penyerta dari domain lain untuk setiap area. Sedangkan model berbasis unit mengasumsikan bahwa variabel penyerta yang tersedia bersesuaian secara individu dengan variabel respon.

Penelitian ini mengaplikasikan model berbasis level area, yakni model Fay-Herriot dengan pertimbangan ketersediaan data pada level unit hanya tersedia pada tahun-tahun pelaksanaan Sensus Penduduk, sehingga sulit untuk melakukan estimasi pada tahun-tahun lainnya.

2.2. Konsep Bayesian pada SAE

Metode Bayesian mengacu nama ilmuwan Thomas Bayes (1702-1761) yang menemukan perlakuan matematika untuk masalah *non trivial* dari inferensi Bayesian. Thomas Bayes menemukan suatu penyelesaian untuk kasus khusus yang kemudian dikenal dengan Teorema Bayesian, selanjutnya dipopulerkan oleh Matematikawan asal Prancis, Peirre-Simon Laplace dengan istilah peluang Bayesian. Dalam imple-

mentasinya, metode Bayes banyak digunakan untuk analisis model statistik yang kompleks (Carlin dan Chib, 1995). Berbeda dengan teori statistik klasik (*frequentist*), analisis bayesian memperlakukan semua parameter yang tidak diketahui sebagai suatu variabel acak dan memiliki distribusi (Hajarisman, 2013).

Secara konseptual, metode Bayesian didasarkan pada teorema Bayes, yaitu 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) dalam bukunya memaparkan jika Y adalah variabel random yang mengikuti pola distribusi tertentu dengan fungsi densitas, $f(\mathbf{y}|\boldsymbol{\theta})$, dengan $\boldsymbol{\theta}$ adalah vektor parameter berukuran d atau $\boldsymbol{\theta}=(\theta_1 \ \theta_2 \ \dots \ \theta_d)^T$ dan $\mathbf{y}=(y_1 \ y_2 \ \dots \ y_n)^T$ adalah vektor sampel berukuran n yang berdistribusi identik dan independen, maka *joint distribution* dari $\boldsymbol{\theta}$ dan \mathbf{y} dapat dituliskan dalam bentuk persamaan :

$$f(\mathbf{y},\boldsymbol{\theta})=f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})=f(\boldsymbol{\theta}|\mathbf{y})f(\mathbf{y}). \quad (2.5)$$

Berdasarkan teorema Bayes, distribusi *posterior* dari $\boldsymbol{\theta}$, $f(\boldsymbol{\theta}|\mathbf{y})$, dapat diturunkan dari persamaan (2.5) sehingga menjadi,

$$f(\boldsymbol{\theta}|\mathbf{y})=\frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y})}, \quad (2.6)$$

dengan $f(\mathbf{y}|\boldsymbol{\theta})$ merupakan fungsi *likelihood* data yang berisi informasi sampel data dan dapat ditulis $f(\mathbf{y}|\boldsymbol{\theta})=\prod_{i=1}^n f(y_i|\boldsymbol{\theta})$. Sedangkan $f(\boldsymbol{\theta})$ adalah fungsi distribusi *prior* dari parameter $\boldsymbol{\theta}$ dan $f(\mathbf{y})$ adalah fungsi konstanta densitas, dimana:

$$f(\mathbf{y})=\begin{cases} \int \dots \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\theta_1 \dots d\theta_d & \text{jika } \theta \text{ kontinu} \\ \sum f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) & \text{jika } \theta \text{ diskrit} \end{cases}$$

Sehingga persamaan (2.6) dapat dinyatakan sebagai berikut:

$$f(\boldsymbol{\theta}|\mathbf{y})\propto f(\mathbf{y}|\boldsymbol{\theta})g(\boldsymbol{\theta}), \quad (2.7)$$

atau bisa juga ditulis, ***Posterior* \propto *Likelihood* \times *Prior***

Fungsi *likelihood* $f(\mathbf{y}|\boldsymbol{\theta})$ memiliki peran yang penting, dimana fungsi tersebut memperbarui informasi tentang *prior* $\boldsymbol{\theta}$, dalam hal ini $f(\boldsymbol{\theta})$, dan dapat dikata-

kan merepresentasikan atau merupakan perwujudan dari informasi yang berasal dari data. Prinsip dari *likelihood* adalah dengan data sampel yang diberikan, untuk setiap dua model peluang $f(y|\theta)$ yang memiliki fungsi *likelihood* yang sama, maka akan menghasilkan inferensia θ yang sama.

Fungsi *likelihood* merupakan representasi dari kondisi data, sedangkan penentuan distribusi *prior* lebih kepada subjektivitas peneliti berdasarkan pertimbangan tertentu. Spesifikasi distribusi *prior* dalam inferensi Bayes juga penting karena distribusi *prior* akan mempengaruhi inferensi mengenai distribusi *posterior*nya (Hajarisman, 2013). Bahkan penentuan distribusi *prior* ini dikatakan sebagai kunci pada inferensi analisis dengan Bayes, sehingga penentuannya menjadi tahapan yang paling penting dalam menggambarkan inferensi.

Setelah distribusi *prior* dispesifikasikan, selanjutnya proses untuk mendapatkan distribusi *posterior* dari fungsi *likelihood* dan distribusi *prior* melibatkan proses analitik ataupun integral numerik yang rumit dan sulit dipecahkan. Dalam metode Bayesian, hal ini dapat diatasi melalui penggunaan *Markov Chain Monte Carlo* (MCMC) (King, Morgan, Gimenez, & Brooks, 2010). Melalui metode MCMC dimungkinkan untuk membangkitkan sampel dari sembarang fungsi densitas *posterior* $f(\theta|y)$ kemudian menggunakan sampel tersebut untuk menghitung nilai harapan dari besaran *posterior* yang akan dikaji.

Satu hal yang penting dalam penggunaan MCMC adalah jika algoritma simulasi diimplementasikan dengan benar, maka rantai Markov akan konvergen ke distribusi target. Implementasi metode MCMC untuk inferensi Bayesian memerlukan algoritma sampling yang tepat untuk mendapatkan sampel dari suatu distribusi. Beberapa algoritma yang dikembangkan untuk proses numerik dalam metode MCMC ini diantaranya adalah algoritma Gibbs Sampling dan *Metropolis-Hasting*.

2.2.1. Metode HB dalam SAE

Melalui pendekatan Bayes, estimasi parameter pada area kecil dapat dilakukan dengan 2 cara yaitu menggunakan *Empirical Bayes* (EB) dan *Hierarchical Bayes* (HB). Untuk pendekatan menggunakan EB, estimasi didasarkan pada distribusi *prior* yang diestimasi dari data, sedangkan pada pendekatan HB parameter model yang ti-

dak diketahui diperlakukan sebagai komponen acak yang memiliki distribusi *prior* tertentu. Pendekatan Bayes, baik EB maupun HB merupakan metode yang dapat diaplikasikan secara luas sehingga dapat digunakan untuk data diskrit, misalkan untuk data biner atau data cacahan.

Pada metode EB, distribusi *prior* dan *posterior* pada area kecil tertentu diestimasi dari data sampel yang ada. Bentuk distribusi bergantung pada parameter model yang tidak diketahui, yang pada umumnya diestimasi melalui *Maximum Likelihood* (ML) dan *Restricted Maximum Likelihood* (REML). Hasil estimasi dimasukkan sebagai informasi pada distribusi *posterior*. Selanjutnya estimasi parameter pada area kecil didasarkan pada distribusi *posterior* marginalnya. Permasalahan mendasar dalam metode EB adalah adanya ketidakpastian dalam estimasi *prior* maupun *posterior*.

Pada metode HB, parameter pengaruh tetap (*fixed effect*) dan komponen varians diperlakukan sebagai sesuatu yang acak. Metode ini mengasumsikan suatu *joint distribution prior* dari suatu parameter, dan teorema Bayes digunakan untuk mendefinisikan *joint distribution prior* dari variabel suatu area kecil. Metode HB biasanya dilakukan dalam beberapa tahapan. Misalnya, Fay dan Herriot (1979) awalnya menurunkan distribusi *posterior* dari β yang ada dan komponen varians σ_v^2 , kemudian menurunkan distribusi *prior* dari σ_v^2 .

Metode HB mempunyai keuntungan karena pemodelan dilakukan secara bertahap, setiap tahap relatif sederhana dan mudah dipahami meskipun proses pemodelannya secara keseluruhan sangat rumit. Orme (2000) dalam *series paper* penelitiannya menyatakan bahwa penggunaan HB jauh lebih baik daripada regresi berganda dalam menghasilkan estimasi parameter yang akurat dan reliabel. Koefisien determinasi pada pemodelan HB lebih tinggi daripada pada pemodelan regresi berganda. Hal ini dikarenakan regresi berganda hanya menghubungkan antara variabel respon dengan variabel independen melalui data sampel secara agregat. Di sisi lain, pemodelan HB jauh lebih baik karena “meminjam” informasi dari observasi yang lain untuk lebih menstabilkan estimasi. Dengan kata lain, pemodelan HB mempertimbangkan pengaruh tiap observasi dalam estimasi.

2.2.2. HB untuk Data Binari

HB menggunakan model *logit-normal* dengan kovariat level area \mathbf{z}_i untuk data binari dapat dituliskan pada persamaan berikut:

- (i) $y_i | p_i \sim \text{Binomial}(n_i, p_i)$
 - (ii) $\theta_i = \text{logit}(p_i) = \mathbf{z}_i^T \boldsymbol{\beta} + v_i$, dengan $v_i \sim N(0, \sigma_v^2)$
 - (iii) $\boldsymbol{\beta}$ dan σ_v^2 saling bebas.
- (2.8)

Misal digunakan *flat prior* untuk $\boldsymbol{\beta}$, $f(\boldsymbol{\beta}) \propto 1$ dan *prior* untuk σ_v^{-2} berdistribusi $\text{Gamma}(a, b)$, dengan $a \geq 0$ dan $b \geq 0$, maka *Gibbs conditional* yang sesuai dengan persamaan (2.3) dapat dituliskan:

- (i) $[\boldsymbol{\beta} | \mathbf{p}_i, \sigma_v^2, \mathbf{y}] \sim N_p \left(\boldsymbol{\beta}, \sigma_v^2 \left(\sum_{i=1}^m \sum_{k=1}^q \mathbf{z}_i \mathbf{z}_i^T \right)^{-1} \right)$
 - (ii) $[\sigma_v^{-2} | \boldsymbol{\beta}, \mathbf{p}, \mathbf{y}] \sim G \left(\frac{m}{2} + a, b + \frac{1}{2} \sum_{i=1}^m (\theta_i - \mathbf{z}_i^T \boldsymbol{\beta})^2 \right)$
 - (iii) $f[p_i | \boldsymbol{\beta}, \sigma_v^2, \mathbf{y}] \propto h[p_i | \boldsymbol{\beta}, \sigma_v^2, \mathbf{y}] k(p_i)$,
- (2.9)

dimana :

$$i = 1, 2, \dots, m,$$

$$\boldsymbol{\beta}^* = \left(\sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^T \right)^{-1} \left(\sum_{i=1}^m \mathbf{z}_i^T \theta_i \right),$$

$$k(p_i) = p_i^{y_i} (1 - p_i)^{n_i - y_i},$$

$$h[p_i | \boldsymbol{\beta}, \sigma_v^2, \mathbf{y}] \propto g(p_i) \exp \left\{ - \frac{(\theta_i - \mathbf{z}_i^T \boldsymbol{\beta})^2}{2\sigma_v^2} \right\},$$

dengan $g(p_i) = \frac{\partial g(p_i)}{\partial p_i}$, dan $p_i = \text{logit}(p_i)$.

Pada persamaan (2.9), hanya *conditional distribution* (i) dan (ii) bersifat *close form*, sehingga digunakan $h[p_i | \boldsymbol{\beta}, \sigma_v^2, \mathbf{y}]$ untuk memperoleh kandidat p_i^* , dengan $p_i = g^{-1}(\theta_i)$ dan $\theta_i | \boldsymbol{\beta}, \sigma_v^2 \sim N(\mathbf{z}_i^T \boldsymbol{\beta}, \sigma_v^2)$. Estimasi model HB untuk p_i dan varians

dari *prior* p_i dihasilkan secara langsung dari sampel MCMC. Estimasi untuk p_i dan varians dari *prior* p_i dapat dituliskan:

$$\hat{p}_i^H \approx \frac{1}{D} \sum_{j=d+1}^{d+D} p_i^{(j)} = p_i^{(\cdot)} \text{ dan } v(p_i | \hat{p}) \approx \frac{1}{D-1} \sum_{j=d+1}^{d+D} (p_i^{(j)} - p_i^{(\cdot)})^2.$$

2.2.3. Distribusi Prior

Spesifikasi distribusi *prior* dalam inferensi *Bayes* memegang peranan yang penting karena akan mempengaruhi inferensi mengenai distribusi *posterior*-nya. Umumnya spesifikasi distribusi *prior* ditekankan pada parameter rata-rata, yang memberikan suatu estimasi titik untuk parameter yang diamati, dan variansnya, yang menyatakan ketidakpastian mengenai nilai estimasi titik tersebut. Apabila secara *apriori* peneliti mempunyai keyakinan yang kuat bahwa estimator tersebut akurat, maka varians seharusnya mempunyai nilai yang kecil, demikian juga sebaliknya.

Terdapat macam-macam *prior*, diantaranya:

1. *Conjugate* dan *Non-Conjugate Prior*. *Conjugate prior* adalah *prior* yang dikaitkan dengan pola model *likelihood* dari datanya (Ntzoufras, 2009). Suatu *prior* dikatakan bersifat *conjugate* untuk keluarga dari distribusi apabila distribusi *prior* dan *posterior*-nya berasal keluarga yang sama, artinya bentuk dari distribusi *posterior* mempunyai bentuk distribusi yang sama sebagai distribusi *prior*. Sebaliknya, apabila suatu *prior* dikatakan sebagai suatu *prior* yang bersifat *non-conjugate* untuk keluarga dari distribusi tertentu apabila distribusi *prior* dan *posterior*-nya bukan berasal keluarga yang sama.
2. *Informative* dan *Non-Informative Prior*. Dikatakan *informative* ataupun *non-informative* dilihat dari sudah diketahui pola dari data observasi atau belum (Ntzoufras, 2009). Banyak peneliti menggunakan distribusi *prior* yang *non-informative* karena distribusi tersebut dianggap lebih objektif. Namun demikian, *prior* yang bersifat *noninformative* ini tidak sepenuhnya dapat menggambarkan total *error* dari parameter yang sedang diamati. Dalam beberapa kasus, *prior* yang *noninformative* dapat membawa pada distribusi *posterior* yang bersifat *improper*, artinya fungsi kepekatan peluang tidak dapat diintegrasikan, yang pada

akhirnya tidak dapat membuat inferensi berdasarkan pada distribusi posterior yang bersifat *improper*.

3. *Proper dan Improper Prior*. Dikatakan *improper* apabila fungsi yang digunakan sebagai “densitas peluang *prior*” memiliki integral yang infinit (tidak dapat diselesaikan) (Ntzoufras, 2009). *Improper* maksudnya adalah distribusinya tidak terintegrasi pada satu. Dengan kata lain, suatu distribusi *prior* disebut *improper* apabila $\int f(\theta) d = \infty$. *Prior* yang bersifat *improper* sering juga digunakan dalam inferensi *Bayes*, karena biasanya dapat menghasilkan *prior* yang bersifat *non-informative* dan distribusi *posterior* yang bersifat *proper*. Namun pada beberapa kasus, distribusi *prior* yang *improper* ini dapat membawa pada distribusi *posterior* yang *improper*. Apabila suatu distribusi *prior* yang *improper* menghasilkan distribusi *posterior* yang *improper*, maka inferensi *Bayes* berdasarkan distribusi *prior* yang *improper* ini menjadi tidak valid.
4. *Pseudo-prior*. *Prior* terkait dengan pemberian nilainya yang disetarakan dengan hasil elaborasi dari frekuentis (Carlin & Chib, 1995).

2.2.4. *Marcov Chain Monte Carlo* (MCMC)

Pada penggunaan metode *Bayesian*, ada kalanya dihadapkan pada kondisi di mana penentuan distribusi *posterior* sulit dilakukan karena melibatkan proses analitik dan integral numerik yang sangat kompleks, misalnya pada model yang kompleks seperti model hirarki dengan banyak parameter, mengakibatkan diperlukannya proses integral dengan dimensi yang besar dan waktu yang cukup lama untuk mendapatkan distribusi *posterior* parameter. Salah satu solusi untuk mengatasi masalah tersebut adalah dengan pendekatan numerik, yaitu MCMC (Carlin dan Chib, 1995).

Pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks. Selain itu, metode ini memungkinkan proses simulasi dengan mengambil sampel random dari model stokastik yang sangat rumit. Ide dasar dari MCMC yaitu membangkitkan data parameter sesuai proses *Marcov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif hingga diperoleh distribusi *posterior* yang stasioner (Ntzoufras, 2009). Dibandingkan dengan teknik simulasi langsung, MCMC lebih bersifat umum dan fleksibel karena simulasi langsung lebih menitikberatkan pada efektifitas penghitungan integral ter-

tentu sehingga tidak dapat digunakan untuk membangkitkan sampel dari berbagai bentuk distribusi *posterior* yang ada.

Misal $\{\theta^{(t)}, t = 1, 2, \dots\}$ merupakan proses stokastik suatu nilai pada periode ke- t dan diinginkan suatu distribusi dari deret pada nilai-nilai berurutan serta diasumsikan bahwa $\theta^{(t)}$ merupakan variabel random yang saling independen. Jika $\theta^{(t)} = i$, maka proses dikatakan dalam keadaan i pada waktu ke t . Dapat dituliskan seperti pada persamaan (2.10):

$$f\left(\theta^{(t+1)} \mid \theta^{(t)}, \dots, \theta^{(1)}\right) = f\left(\theta^{(t+1)} \mid \theta^{(t)}\right) \quad (2.10)$$

Proses stokastik ini dikenal sebagai *Markov Chain* (Ross, 2014). Interpretasi dari persamaan (2.10) yaitu distribusi bersyarat untuk semua keadaan $\theta^{(t+1)}$ dengan syarat kondisi sebelumnya $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(t-1)}$ dan keadaan saat ini $\theta^{(t)}$ bersifat independen dengan keadaan sebelumnya, dan hanya bergantung pada keadaan saat ini. Pada $t \rightarrow \infty$, distribusi dari $\theta^{(t)}$ akan konvergen menuju distribusi tertentu yang independen terhadap nilai awal dari rantai tersebut (Ntzoufras, 2009).

Kondisi *equilibrium* tercapai jika sampel yang diperoleh telah memenuhi sifat dari *Markov chain*, yaitu:

- a. *Irreducible*, yaitu selama iterasi parameter dalam proses MCMC memiliki perubahan nilai yang acak sebagai gambaran dari sifat *communicate* antar keadaan dalam *Markov chain*.
- b. *Aperiodic*, yaitu selama iterasi dalam proses MCMC tidak memiliki periode tertentu yang pasti kembali ke keadaan semula, karena nilai keadaan *Markov chain* yang diperoleh bersifat kontinu sehingga kemungkinan kecil untuk mendapatkan nilai yang sama antara iterasi satu dengan yang lain, dan
- c. *Recurrent*, yaitu selama proses iterasi dalam MCMC, terdapat adanya kepastian bahwa nilai parameter yang dibangkitkan akan muncul kembali pada proses iterasi yang sangat panjang berikutnya.

Metode MCMC menggabungkan prosedur iteratif, karena nilai yang ada pada setiap langkah bergantung pada nilai iterasi sebelumnya. Menurut Ntzoufras (2009), algoritma dari MCMC adalah sebagai berikut:

1. Tentukan nilai awal untuk tiap parameter model dengan memperhatikan karakteristik datanya.
2. Bangkitkan C sampel $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(P)}\}$ dari distribusi *posterior* $f(\boldsymbol{\theta}|\mathbf{y})$ secara *full conditional*.
3. Amati konvergensi data sampel yang dihasilkan dari proses iterasi. Jika konvergensi belum tercapai, maka periksa kembali domain *prior*, nilai awal, atau diperlukan sampel lebih banyak lagi. Ulangi langkah 2 hingga konvergensi tercapai.
4. Lakukan proses *burn in*. Periode *burn in* merupakan periode awal iterasi estimasi parameter dalam proses MCMC dengan membuang sebanyak A iterasi dalam rangka menghilangkan pengaruh dari penggunaan nilai awal (*initial value*). Periode *burn in* akan berakhir ketika kondisi *equilibrium* tercapai.
5. Gunakan $\{\boldsymbol{\theta}^{(B+1)}, \boldsymbol{\theta}^{(B+2)}, \dots, \boldsymbol{\theta}^{(P)}\}$ sebagai sampel dalam analisis distribusi *posterior*.
6. Buat plot distribusi *posterior*.
7. Buat ringkasan statistik dari distribusi *prior* (rata-rata, median, standar deviasi, kuantil, korelasi, MC error dan 95% *credible interval*).

Untuk mengetahui variasi dari hasil estimasi *prior* parameter yang diperoleh dari proses simulasi MCMC digunakan ukuran *Monte Carlo error* (MC error). Nilai MC error yang kecil mengindikasikan akurasi yang tinggi dari estimasi *prior* yang diperoleh (Ntzoufras, 2009).

Proses diagnostik yang perlu dilakukan yaitu diagnostik konvergensi rantai Markov dan diagnostik untuk mengevaluasi kecocokan model. Pemeriksaan konvergensi dilakukan melalui pendekatan visualisasi *trace plot* sampel dengan indeks iterasi. *Trace* ini mendeskripsikan tentang apakah rantai sudah mencapai kekonvergenan terhadap suatu distribusi yang stasioner atau belum. Apabila belum mencapai kekonvergenan, biasanya periode *burn-in* perlu diperpanjang. Suatu rantai dikatakan sudah stasioner apabila distribusi dari titik-titik tidak berubah. Konsep kestasioneran di sini dapat dilihat dari *trace plot* apabila rata-rata dan ragamnya relatif konstan. *Density plot* dan plot autokorelasi juga menunjukkan konvergensi apabila hasil *density* menunjukkan pola yang *smooth* dan plot autokorelasi menurun.

2.2.5. Gibbs Sampling

Implementasi metode MCMC untuk analisis Bayesian memerlukan algoritma sampling yang tepat untuk mendapatkan sampel dari suatu distribusi. Algoritma yang sering digunakan sebagai pembangkit variabel random dalam MCMC adalah *Gibbs Sampling* (Gelman dkk, 2014). *Gibbs sampling* dapat didefinisikan sebagai suatu teknik simulasi untuk membangkitkan variabel random dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella dan George, 1992).

Gibbs sampler merupakan generator yang sangat efisien sehingga sering digunakan sebagai generator variabel random pada analisis data yang menggunakan MCMC. Proses ini dilakukan dengan mengambil sampel dengan cara membangkitkan rangkaian *gibbs* variabel random berdasarkan sifat-sifat dasar proses Markov Chain. Dalam menjalankan program yang menggunakan rantai markov dilakukan pada kondisi bersyarat penuh. Ini merupakan salah satu kelebihan dari *Gibbs Sampling* karena variabel random tersebut dibangkitkan dengan menggunakan konsep distribusi unidimensional yang terstruktur sebagai distribusi *full conditional*. *Gibbs sampling* sangat berguna dalam mengestimasi suatu parameter dalam suatu model kompleks yang mempunyai tingkat kerumitan dalam proses integritasi yang kompleks pula dan sulit diselesaikan secara analitis.

Ntzoufras (2009) menjelaskan algoritma Gibbs Sampling sebagai berikut:

1. Tetapkan nilai awal parameter θ pada $t = 0$ sehingga $\theta^{(0)} = (\theta_1^{(0)} \quad \dots \quad \theta_d^{(0)})^T$
2. Untuk $t = 1, \dots, M$, ulangi langkah:
 - a. Tentukan $\theta = \theta^{(t-1)}$.
 - b. Untuk $s = 1, 2, \dots, d$, update θ_s dari $\theta_s \sim f(\theta_s | \theta_{\setminus s}, \mathbf{y})$.
 - c. Tentukan $\theta^{(t)} = \theta$ dan gunakan untuk membangkitkan iterasi ke $t + 1$.

Berikut adalah proses *sampling* untuk mendapatkan nilai $\theta^{(t)}$,

$$\begin{aligned} &\theta_1^{(t)} \text{ dari } f(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}, \mathbf{y}), \\ &\theta_2^{(t)} \text{ dari } f(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_d^{(t-1)}, \mathbf{y}), \\ &\vdots \\ &\theta_s^{(t)} \text{ dari } f(\theta_s | \theta_1^{(t)}, \theta_2^{(t-1)}, \dots, \theta_{s-1}^{(t-1)}, \theta_{s+1}^{(t-1)}, \theta_d^{(t-1)}, \mathbf{y}), \end{aligned}$$

⋮

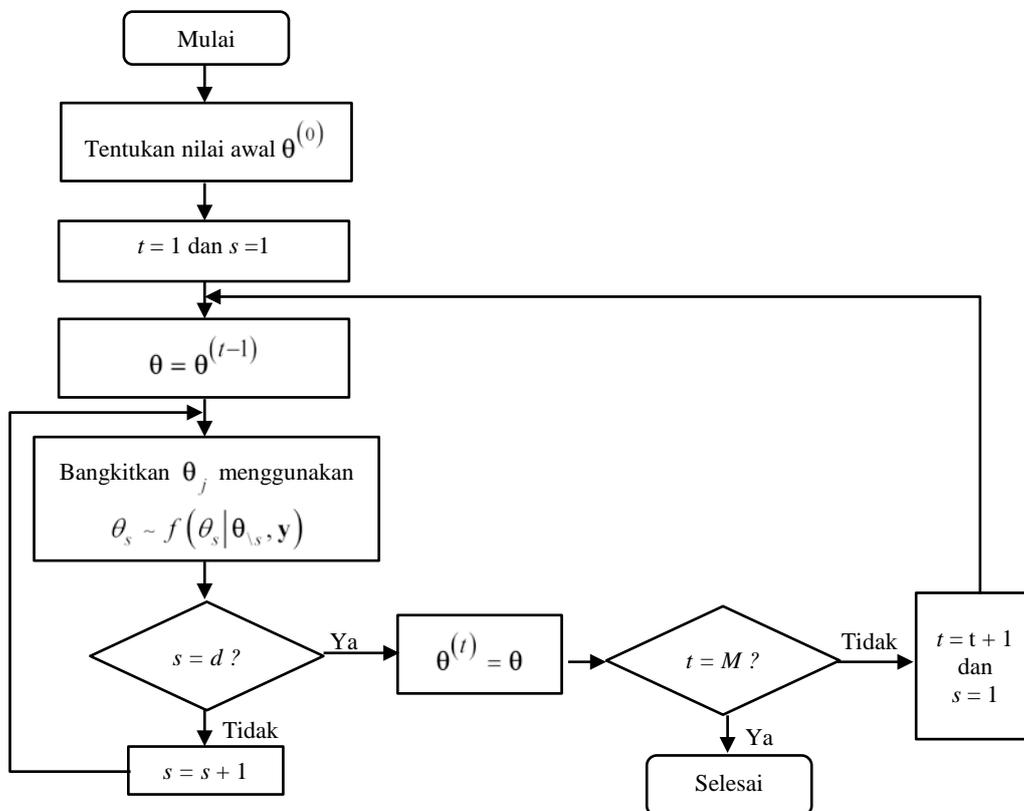
$\theta_d^{(t)}$ dari $f(\theta_d | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{d-1}^{(t)}, \mathbf{y})$.

Pembangkitan nilai dari $f(\theta_s | \theta_{\setminus s}, \mathbf{y}) = f(\theta_s | \theta_1^{(t)}, \theta_2^{(t-1)}, \dots, \theta_{s-1}^{(t-1)}, \theta_{s+1}^{(t-1)}, \theta_d^{(t-1)}, \mathbf{y})$

adalah relatif mudah karena merupakan distribusi univariat dan dapat ditulis sebagai

$f(\theta_s | \theta_{\setminus s}, \mathbf{y}) \propto f(\boldsymbol{\theta} | \mathbf{y})$, dimana semua variabel lain, kecuali θ_j , adalah konstan.

Algoritma *Gibbs Sampling* dijelaskan dengan lebih sederhana melalui diagram alur pada Gambar 2.1. (Ismartini, 2013).



Gambar 2.1. Diagram Alur Algoritma *Gibbs Sampling*

2.3. Distribusi Multinomial

Distribusi Multinomial, generalisasi multivariat dari distribusi binomial, digunakan untuk menemukan probabilitas dari percobaan di mana ada lebih dari dua hasil. Misalkan setiap percobaan dapat menghasilkan *output* E_1, E_2, \dots, E_k dengan masing-masing probabilitas p_1, p_2, \dots, p_k . Distribusi multinomial akan memberikan

probabilitas bahwa E_1 akan muncul sebanyak y_1 kali, E_2 akan muncul sebanyak y_2 kali, dan seterusnya, dalam pengambilan independen sebanyak n kali. Distribusi multinomial mempunyai bentuk *Probability Mass Function* (PMF) sebagai berikut :

$$f(y_1, \dots, y_k, p_1, \dots, p_k, n) = \frac{n!}{y_1! y_2! \dots y_k!} (p_1^{y_1}) (p_2^{y_2}) \dots (p_k^{y_k})$$

$$f(y|\theta, \phi) = \frac{n!}{y_1! y_2! \dots y_k!} (p_1^{y_1}) (p_2^{y_2}) \dots (p_k^{y_k}) \quad (2.11)$$

Jika PMF dari distribusi Multinomial di-*log* lalu diekspensialkan kedua sisinya, maka diperoleh:

$$\begin{aligned} f(y|\theta, \phi) &= \frac{n!}{y_1! y_2! \dots y_k!} (p_1^{y_1}) (p_2^{y_2}) \dots (p_k^{y_k}) \\ &= \exp \left(\log \left(\left(\frac{n!}{y_1! y_2! \dots y_k!} \right) (p_1^{y_1}) (p_2^{y_2}) \dots (p_k^{y_k}) \right) \right) \\ &= \exp \left(\log \left(\frac{n!}{y_1! y_2! \dots y_k!} \right) + \log(p_1^{y_1}) + \log(p_2^{y_2}) + \dots + \log(p_k^{y_k}) \right) \\ &= \exp \left(\log \left(\frac{n!}{y_1! y_2! \dots y_k!} \right) + y_1 \log(p_1) + y_2 \log(p_2) + \dots + y_k \log(p_k) \right) \\ &= \exp \left(y_1 \log(p_1) + y_2 \log(p_2) + \dots + y_k \log(p_k) + \log \left(\frac{n!}{y_1! y_2! \dots y_k!} \right) \right) \\ &= \exp \left(\sum_{j=1}^k y_j \log p_j + \log \left(\frac{n!}{y_1! y_2! \dots y_k!} \right) \right) \\ &= \exp \left(\sum_{j=1}^{k-1} y_j \log p_j + \left(1 - \sum_{j=1}^{k-1} y_j \right) \left(1 - \sum_{j=1}^{k-1} p_j \right) \log \left(\frac{n!}{y_1! y_2! \dots y_k!} \right) \right) \end{aligned}$$

Dengan melihat konsep bentuk fungsi keluarga eksponensial, maka dapat dilakukan pemisalan, sehingga diperoleh nilai θ sebagai berikut:

misal, $\theta = \log \frac{p_k}{\left(1 - \sum_{j=1}^{k-1} p_j \right)}$

$$\Rightarrow e^\theta = \frac{p}{1-p}$$

$$\Rightarrow p = \frac{e^\theta}{1+e^\theta}$$

$$\Rightarrow 1-p = \frac{1+e^\theta - e^\theta}{1+e^\theta} = (1+e^\theta)^{-1}$$

$$\Rightarrow n \log(1-p) = n \log(1+e^\theta)^{-1} = n \log(1+e^\theta)$$

maka,

$$f(y|\boldsymbol{\theta}, \phi) = \exp\left(\theta y - n \log(1+e^\theta) + \log\binom{n}{y}\right)$$

Sehingga, $\theta = \log\left(\frac{p_k}{1 - \sum_{j=1}^{k-1} p_j}\right)$, $\phi = 1$,

$$b(\theta) = n \log(1+e^\theta), \quad a(\phi) = 1, \quad c(y, \phi) = \log\left(\frac{n!}{y_1! y_2! \dots y_k}\right),$$

$$\begin{aligned} \text{Mean } (\mu) \quad : \quad E(Y) &= \frac{\partial}{\partial \theta} b(\theta) = \frac{\partial}{\partial \theta} (n \log(1+e^\theta)) \\ &= n \frac{e^\theta}{1+e^\theta} = np \end{aligned} \tag{2.12}$$

$$\begin{aligned} \text{Varians} \quad : \quad \text{Var}(Y) &= \frac{\partial^2}{\partial \theta^2} b(\theta) a(\phi) = \frac{\partial}{\partial \theta} \left(n \frac{e^\theta}{1+e^\theta} \right) 1 \\ &= n \frac{e^\theta}{(1+e^\theta)^2} = n \frac{e^\theta}{(1+e^\theta)} \frac{1}{(1+e^\theta)} \\ &= np(1-p) \end{aligned} \tag{2.13}$$

$$\text{Link Function} \quad : \quad \eta = \theta = \log\left(\frac{p_k}{1 - \sum_{j=1}^{k-1} p_j}\right). \tag{2.14}$$

2.4. WinBUGS

WinBUGS adalah *software* berbasis bahasa pemrograman yang digunakan untuk menghasilkan sampel acak dari distribusi posterior parameter model Bayesian (Ntzoufras, 2009). User hanya perlu menentukan data, struktur model, dan beberapa nilai awal parameter model. WinBUGS merupakan perluasan dari BUGS (*Bayesian Inference Using Gibbs Sampling*) yang didesain berbasis *windows* dan merupakan *open source software*.

Tujuan awal dari WinBUGS adalah untuk mengembangkan *software* yang dapat menghasilkan distribusi *posterior* parameter model menggunakan MCMC. Model tersebut dapat dispesifikasi dalam WinBUGS menggunakan code yang relatif sederhana. Code dalam WinBUGS mirip dengan bahasa pemrograman yang digunakan dalam *software* Splus dan R. WinBUGS juga menyediakan kemudahan dalam membangun code yaitu melalui menu *Doodle interface*. Menu ini berguna untuk mengidentifikasi struktur model dengan menggambar struktur grafis, biasa dikenal dengan *Directed Acyclic Graph* (DAG). WinBUGS ini secara otomatis menghasilkan code model sesuai struktur grafis DAG yang telah dibuat. DAG disusun dengan dua *node* yang berbeda, yaitu kotak dan oval. *Node* kotak menunjukkan nilai parameter yang konstan dan *node* oval menunjukkan komponen stokastik dalam model.

2.5. Koefisien Variasi

Koefisien variasi (*coefficient of variance/CV*) merupakan ukuran kekonvergenan dari estimasi yang dihasilkan (Zhou dan You, 2008). Nilai CV dari estimasi distribusi *prior* hasil estimasi model HB akan dibandingkan dengan nilai CV dari estimasi langsung. Nilai CV diperoleh dengan membagi akar kuadrat varians dengan rata-rata atau proporsi, atau dapat ditulis sebagai berikut:

$$CV = \frac{\sqrt{V}}{m}. \quad (2.15)$$

Untuk estimasi langsung, Cochran (1977) merumuskan bahwa *unbiased* varians untuk estimasi proporsi yang berasal dari sampel dapat dituliskan:

$$\bar{V}(\hat{p}) = \frac{N-n}{N} \times \frac{\hat{p}(1-\hat{p})}{n-1}. \quad (2.16)$$

Sedangkan untuk rumus proporsi digunakan rumus:

$$\hat{p} = \frac{\sum_{i=1}^n y_i}{n}, \quad (2.17)$$

dimana \hat{p} adalah proporsi yang diperoleh dari estimasi sampel, N jumlah populasi, dan n jumlah sampel yang diambil.

2.6. Estimasi Mean Square Error (MSE) Jackknife

Menurut Baíllo dan Molina (2009), tujuan dari prosedur dan teknik yang digunakan dalam SAE adalah untuk memperoleh estimasi dengan tingkat presisi yang tinggi pada area kecil tersebut. Tingkat presisi estimator ini dapat digambarkan oleh MSE. Rao (2007) menyatakan bahwa untuk SAE dengan pendekatan EB pada model Fay-Herriot, MSE dari dapat dituliskan seperti persamaan (2.18):

$$\begin{aligned} MSE(\hat{\theta}_i^E) &= E(\hat{\theta}_i^E - \theta_i)^2 \\ &= E(\hat{\theta}_i^B - \theta_i)^2 + E(\hat{\theta}_i^E - \theta_i)^2 \\ &= \hat{M}_{1i}(\sigma_v^2) + \hat{M}_{2i}(\sigma_v^2) \end{aligned} \quad (2.18)$$

dimana $i = 1, 2, \dots, m$, $\hat{\theta}_i^B$ merupakan estimator terbaik dari θ_i . $\hat{\theta}_i^E$ diperoleh dari substitusi $\hat{\beta}$ dan $\hat{\sigma}_v^2$ pada $\hat{\theta}_i^B$.

Rao (2007) menjelaskan bahwa ketepatan model dalam SAE dengan kriteria bias sangat sulit dilakukan karena nilai parameter populasi pada wilayah kecil tidak diketahui, sehingga parameter populasi juga diestimasi melalui estimasi dari sampel yang tersedia. $\hat{M}_{1i}(\sigma_v^2)$ pada persamaan (2.18) sama dengan $g_{1i}(\sigma_v^2)$ yang menunjukkan efisiensi dari estimator $\hat{\theta}_i$. Prasad dan Rao (1990), Datta dan Lahiri (2000) serta Datta, Rao, dan Smith (2005) dalam Rao (2007) menggunakan *Taylor linearization* untuk mengestimasi MSE yaitu untuk mengoreksi ketidakpastian akibat menduga parameter populasi menggunakan data sampel. *Taylor linearization* digunakan untuk pendekatan pada $\hat{M}_{2i}(\sigma_v^2)$ pada persamaan (2.18) untuk m besar yaitu $\hat{M}_{2i}(\sigma_v^2) \approx g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2)$. Namun, Rao juga menyebutkan bahwa menduga MSE menggunakan *Taylor linearization* merupakan sesuatu yang kompleks dan sulit. Estimator MSE menggunakan *Taylor linearization* dituliskan sebagai berikut:

$$MSE(\hat{\theta}_i^E) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2) \quad (2.19)$$

Rao menyatakan bahwa Jiang, Lahiri, dan Wan (2002) mengusulkan metode *Jackknife* untuk mengoreksi $g_{2i}(\cdot)$ dan $g_{3i}(\cdot)$ pada persamaan (2.19). Jiang, Lahiri, dan Wan mengaplikasikan *Jackknife* yang dikembangkan oleh Turkey (1958). Metode *Jackknife* merupakan metode untuk mengoreksi bias dari suatu penduga. Metode ini merupakan metode resampling yang dilakukan dengan membangkitkan data yang berasal dari sampel sehingga akan mendekati parameter populasinya. Penerapan *Jackknife* pada SAE dilakukan untuk mengoreksi pendugaan MSE. Estimator MSE *Jackknife* tak berbobot Jiang, Lahiri dan Wan (JLW) dari $\hat{M}_{2i}(\hat{\sigma}_v^2)$ pada persamaan (2.18) dapat dituliskan sebagai berikut:

$$\hat{M}_{2ij} = \frac{m-1}{m} \sum_{l=1}^m \left(h(\hat{\theta}_i, \hat{\phi}_{(l)}) - h(\hat{\theta}_i, \hat{\phi}) \right)^2.$$

Pada kasus untuk menghitung proporsi, \hat{M}_{2ij} dapat dituliskan sebagai berikut:

$$\hat{M}_{2ij} = \frac{m-1}{m} \sum_{l=1}^m \left(\hat{p}_{i(l)} - \hat{p}_i \right)^2,$$

dimana $\hat{p}_{i(l)}$ merupakan \hat{p}_i dengan menghilangkan data ke- l , $l = 1, 2, \dots, m$. Reduksi bias diaplikasikan ke $\hat{M}_{1ij}(\hat{\sigma}_v^2)$:

$$\hat{M}_{1ij} = g_{1i}(\hat{\sigma}_v^2) - \frac{m-1}{m} \sum_{l=1}^m \left[g_{1i}(\hat{\sigma}_{v(l)}^2) - g_{1i}(\hat{\sigma}_v^2) \right],$$

dengan j replikasi pada prosedur *Jackknife*, $g_{1i}(\hat{\sigma}_{v(l)}^2)$ merupakan $g_{1i}(\hat{\sigma}_v^2)$ dengan menghilangkan data ke- l , $l = 1, 2, \dots, m$.

Rao menyatakan bahwa metode *Jackknife* yang dikembangkan oleh Jiang, Lahiri, dan Wan dapat digunakan untuk semua model untuk SAE, termasuk juga *mismatched model* dan untuk kasus yang tidak berdistribusi normal (data binari atau cacahan).

2.7. Rata-rata Lama Sekolah (RLS)

Estimasi langsung dapat digunakan apabila semua area dalam populasi digunakan sebagai sampel dan estimator ini berbasis desain sampling (Rao, 2003). Esti-

masi langsung menggunakan nilai dari variabel yang menjadi perhatian hanya pada periode waktu dan unit sampel area (Ramsini dkk dalam Fausi dan Sutikno, 2011). Data sampel dari suatu survei dapat digunakan untuk mendapatkan pendugaan langsung yang dapat dipercaya bagi suatu area besar atau domain. Nilai hasil pendugaan langsung pada suatu area kecil merupakan estimator tak bias meskipun memiliki varian yang besar dikarenakan dugaannya diperoleh dari ukuran sampel yang kecil (Pfefferman (2002) dalam Rumiati (2012)).

Estimasi RLS penduduk usia 25 tahun ke atas dapat dirumuskan:

$$LS = \frac{1}{P_{25+}} \sum_{i=1}^{P_{25+}} (\text{Lama sekolah penduduk ke } -i), \quad (2.20)$$

dimana:

P_{25+} = jumlah penduduk berusia 25 tahun ke atas

Lama sekolah penduduk ke- i =

- a. Tidak pernah sekolah = 0
- b. Masih sekolah di SD sampai dengan S1 = konversi ijazah terakhir + kelas terakhir - 1
- c. Masih sekolah di S2/S3 = konversi ijazah terakhir + 1
- d. Tidak bersekolah lagi dan tamat di kelas terakhir = konversi ijazah terakhir
- e. Tidak bersekolah lagi dan tidak tamat di kelas terakhir = konversi ijazah terakhir + kelas terakhir - 1.

Penduduk yang tamat SD diperhitungkan lama sekolah selama 6 tahun, tamat SMP diperhitungkan lama sekolah selama 9 tahun, tamat SMA diperhitungkan lama sekolah selama 12 tahun, tanpa memperhitungkan apakah pernah tinggal kelas atau tidak.

2.8. Survei Sosial Ekonomi Nasional (Susenas) Tahun 2015

Susenas merupakan survei berbasis rumah tangga (ruta) yang dirancang untuk mengumpulkan data sosial kependudukan yang relatif sangat luas terutama terkait dengan tingkat kesejahteraan masyarakat. Susenas dilaksanakan tiap tahun oleh BPS sejak tahun 1963. Indikator yang diperoleh dari Susenas mencakup berbagai bidang antara lain bidang kesehatan, pendidikan, ketenagakerjaan, fertilitas, perumahan, konsumsi/pengeluaran rumah tangga dan sosial ekonomi lainnya.

Susenas 2015 dilaksanakan pada seluruh provinsi di Indonesia. Pencacahan terhadap sampel terpilih dilakukan oleh tim yang terdiri dari pengawas dan pewawancara dan selanjutnya proses wawancara dilakukan terhadap ruta terpilih dan responden terpilih dengan menggunakan kuesioner. Kuesioner Susenas 2015 terdiri dari kuesioner Kor (VSEN16.K) yang berisi sekumpulan pertanyaan umum dan selalu ditanyakan setiap tahun, kuesioner Modul Konsumsi (VSEN16.M) mencakup pertanyaan rinci mengenai kegiatan sosial kemasyarakatan dan kondisi sosial budaya, seperti gangguan disabilitas, akses media massa, rincian besar biaya dan beasiswa pendidikan, partisipasi terhadap organisasi kemasyarakatan dan modal sosial.

Secara umum tujuan pengumpulan data melalui Susenas adalah menyediakan data kesejahteraan rakyat yang mencakup pendidikan, kesehatan, dan kemampuan daya beli. Secara khusus tujuan dari Susenas adalah:

1. Tersedianya data pokok tentang kesejahteraan masyarakat yang dibutuhkan untuk perencanaan, monitoring dan evaluasi keberhasilan pembangunan.
2. Tersedianya data rinci tentang kesejahteraan ART.

Kuesioner Susenas dibagi menjadi dua jenis yaitu KOR, merupakan kumpulan pertanyaan yang bersifat umum, dan Modul, merupakan pertanyaan bersifat lebih spesifik sesuai dengan jenis paket yang dilakukan pada tahun tersebut. Modul Susenas dibagi menjadi tiga paket yang dikumpulkan secara bergiliran dalam kurun waktu tiga tahun. Ketiga paket tersebut adalah (1) konsumsi/pengeluaran dan pendapatan rumah tangga; (2) Sosial Budaya; dan (3) Kesehatan dan perumahan.

Pengumpulan data Susenas tahun 2015 dilaksanakan secara triwulan (tw), yaitu dilaksanakan pada bulan Maret (tw. I), Juni (tw. II), September (tw. III) dan Desember (tw. IV). Pelaksanaan Susenas tahun 2015 mencakup 300.000 sampel ruta yang tersebar di 34 provinsi dan 497 kabupaten/kota di Indonesia, setiap triwulan akan didistribusikan sebanyak 7.500 sampel blok sensus (BS) atau 75.000 sampel ruta (BPS,2016). Data hasil pencacahan tiap triwulan digunakan untuk estimasi level provinsi sedangkan data gabungan hasil pencacahan seluruh triwulan digunakan untuk estimasi sampai level kabupaten/kota.

2.9. Penelitian Terkait Penentuan Variabel Bebas

Suendra (1999) dalam penelitiannya menyatakan bahwa ada beberapa faktor yang menghambat seorang anak untuk menyelesaikan pendidikan dasar 9 tahun, antara lain hambatan sosial budaya, sosial ekonomi, geografis, kurangnya fasilitas pendidikan, dan tenaga kependidikan.

Sukarno (2002) mengangkat isu lama sekolah penduduk usia 15-29 tahun di Jakarta, Kota Semarang dan Kota Surabaya. Penelitian tersebut mengungkapkan bahwa anak laki-laki dengan pendidikan orang tua lebih tinggi, akan memiliki tingkat pencapaian pendidikan yang lebih tinggi pula jika dibandingkan dengan anak perempuan. Pekerjaan bapak juga berpengaruh cukup signifikan terhadap tingkat pencapaian pendidikan, tetapi pengaruhnya kecil. Selain pengaruh dari faktor orang tua dan jenis kelamin, umur, status pernikahan, tempat lahir, dan suku juga mempengaruhi tingkat pencapaian pendidikan. Diungkapkan juga bahwa anak laki-laki yang berstatus belum menikah dan lahir di daerah perkotaan serta bersuku Jawa, selain cenderung mencapai tingkat pendidikan lebih tinggi juga cenderung lebih banyak yang memiliki pendidikan nonformal.

Santoso (2009) mengidentifikasi variabel yang berpengaruh terhadap lama sekolah di Provinsi Papua. Sesuai dengan urutan kuatnya pengaruh, variabel tersebut antara lain topografi wilayah, umur anak, status pekerjaan anak, pendidikan kepala rumahtangga, pengeluaran perkapita, banyaknya anggota rumahtangga, lapangan pekerjaan kepala rumahtangga, status desa/kota dan klasifikasi daerah sulit. Topografi wilayah, status desa/kota dan klasifikasi daerah sulit memberikan gambaran bahwa karakteristik area dapat mempengaruhi lama sekolah penduduknya. Dalam hal gender, kadangkala ketika dukungan ekonomi keluarga tidak memadai untuk membiayai pendidikan anak-anak mereka, maka orangtua akan memilih anak mana yang harus sekolah, anak laki-laki atau perempuan yang bersekolah. Hal ini sangat tergantung pada cara pandang orangtua terhadap nilai anak, termasuk faktor kultural yang melekat pada diri mereka yang dapat mempengaruhi cara pandang tersebut.

Dengan menggunakan model Spasial CAR, Solikhah (2009) mengidentifikasi variabel yang memengaruhi RLS di Pulau Kalimantan yaitu: persentase desa/kelurahan yang memiliki SLTP, persentase desa/kelurahan yang memiliki SLTA, persentase kecamatan yang memiliki Perguruan Tinggi, persentase penduduk muda, persen-

tase penduduk yang tinggal di daerah perkotaan, rasio murid-guru SLTP, rasio murid-guru SLTA, kontribusi sektor non pertanian dalam PDRB, rata-rata pendapatan per kapita 1 bulan, dan rasio jenis kelamin. Menurut teori ekonomi dalam Solikhah (2009), dukungan dan dorongan untuk meneruskan pendidikan bervariasi menurut daya beli keluarga serta akses lembaga pendidikan serta peluang kerja yang tersedia.

Berdasarkan penjelasan di atas pada penelitian ini variabel-variabel prediktor yang diduga mempengaruhi lama sekolah adalah topografi wilayah, jumlah sekolah, dan jarak kecamatan ke ibukota kabupaten.

BAB III

METODOLOGI PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS), yaitu untuk variabel respon berasal dari data hasil Survei Sosial Ekonomi Nasional (Susenas) Tahun 2015. Sedangkan data spesifik berkenaan dengan area kecil berasal dari hasil Pendataan Potensi Desa (2014), dan Kecamatan Dalam Angka yang diterbitkan oleh BPS Kabupaten Keerom. Unit observasi yang digunakan dalam penelitian ini adalah 7 kecamatan di Kabupaten Keerom. Pada penelitian ini, unit observasi yang digunakan dalam penulisan ini menggunakan data Susenas yaitu sebanyak 662 individu sampel Susenas di Kabupaten Keerom pada tahun 2015.

3.2. Metode Pengumpulan Data

Susenas merupakan survei lingkup nasional yang dilaksanakan oleh BPS dengan unit sampel rumah tangga. Susenas 2015 menggunakan tiga jenis kerangka sampel dengan urutan tahapan yaitu kerangka sampel untuk pemilihan kecamatan, kerangka sampel untuk pemilihan blok sensus dari kecamatan terpilih, dan kerangka sampel untuk pemilihan sampel rumah tangga dari blok sensus terpilih.

Setiap rumah tangga yang terpilih dalam survei ini dikunjungi oleh petugas pencacah yang bertugas mewawancarai responden sesuai dengan daftar pertanyaan dalam kuisisioner yang sudah disiapkan. Wawancara dilakukan langsung terhadap kepala rumah tangga atau kepada anggota rumah tangga yang dianggap mengetahui keadaan di rumah tangga tersebut. Referensi waktu survei yang digunakan adalah selama seminggu yang lalu untuk konsumsi makanan, serta satu bulan sampai tiga bulan yang lalu untuk konsumsi non-makanan. Petugas yang melakukan pengumpulan data adalah Koordinator Statistik Kecamatan (KSK) atau staf BPS kabupaten/kota yang sudah dilatih terlebih dahulu dengan konsep dan definisi serta simulasi survei. Setiap petugas survei juga dibekali dengan instrumen untuk keperluan pengumpulan data, seperti kuisisioner dan buku pedoman.

Adapun jumlah sampel terpilih dalam Susenas tahun 2015 di Kabupaten Keerom adalah sebanyak 360 sampel rumah tangga atau 2,76 % dari jumlah rumah tangga di Kabupaten Keerom tahun 2015. Dari sampel rumah tangga tersebut, terdapat 662 penduduk yang berusia 25 tahun ke atas, atau 2,36% dari jumlah penduduk yang berusia 25 tahun ke atas di Kabupaten Keerom pada tahun 2015.

3.3. Variabel Penelitian

Variabel respon yang digunakan dalam penelitian ini adalah lama sekolah penduduk usia 25 tahun ke atas pada tahun 2015, yaitu jumlah tahun belajar penduduk pada usia tersebut (yang tidak pernah menempuh sekolah, yang sedang bersekolah, dan yang pernah menempuh sekolah namun kini sudah tidak bersekolah lagi) yang telah diselesaikan dalam pendidikan formal (tidak termasuk tahun yang mengulang) dalam satuan tahun.

Penelitian ini menerapkan model Multinomial untuk mendapatkan lama sekolah berdasarkan kecamatan dan kategori lama sekolah. Variabel respon yang digunakan pada model Multinomial proporsi lama sekolah menurut kategori lama sekolah di setiap kecamatan, terdiri dari 6 kategori, yaitu:

Kategori 1 : lama sekolah 0 tahun

Kategori 2 : lama sekolah 1-3 tahun

Kategori 3 : lama sekolah 4-6 tahun

Kategori 4 : lama sekolah 7-9 tahun

Kategori 5 : lama sekolah 10-12 tahun

Kategori 6 : lama sekolah 13 tahun ke atas

p_{ij} adalah proporsi lama sekolah berdasarkan kategori lama sekolah pada setiap kecamatan. y_{ij} menunjukkan jumlah penduduk usia 25 tahun ke atas di kecamatan i , $i=1,2,\dots,m$, pada kategori j , $j=1,2,\dots,k$, dimana y_{ij} merupakan data biner yang bernilai 1 untuk penduduk usia 25 tahun ke atas dari kategori 1, dan bernilai nol untuk kategori lainnya, dan seterusnya. N_i adalah jumlah penduduk usia 25 tahun ke atas di kecamatan ke- i .

$$p_{ij} = \frac{\sum_{j=1}^k y_{ij}}{N_i} = \frac{y_i}{N_i}$$

Variabel penyerta yang digunakan pada penelitian ini terdiri dari:

1. **Topografi wilayah** (X_1) merupakan pembagian wilayah berdasarkan dua jenis topografi yaitu lereng/lembah dan dataran, bernilai 1 jika topografi wilayah berupa dataran, dan 0 jika lainnya.
2. **Jumlah SD** (X_2) yaitu jumlah fasilitas pendidikan SD/Sederajat
3. **Jumlah SMP** (X_3) yaitu jumlah fasilitas pendidikan SMP/Sederajat.
4. **Jumlah SMA** (X_4) yaitu jumlah fasilitas pendidikan SMA/Sederajat.
5. **Jarak kecamatan ke ibukota kabupaten** (X_5) yaitu jarak (dalam kilometer) antara kantor kecamatan ke kantor kabupaten.

3.4. Metode untuk Kajian Teori

Tujuan pertama dari penelitian ini yaitu untuk memperoleh estimasi parameter dari *Multinomial Logit Mixed Model* dengan pendekatan *Hierarchical Bayes* yang akan dilakukan dengan langkah-langkah sebagai berikut (Ayuningtyas, 2017):

1. Mendefinisikan model SAE untuk respon multinomial dengan menggunakan *Multinomial Logit Mixed Model*, yaitu $\theta_{ij} = \mathbf{z}_{ij}^T \boldsymbol{\beta}_j + v_{ij}$, dengan *link function* seperti persamaan 2.15, dimana $j = 1, 2, \dots, k, v_i \sim M(0, \mathbf{W}), \boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$.
2. Membentuk kerangka HB model SAE untuk model *linking* dengan model berbasis level area, sehingga diperoleh model *logit normal* dengan kovariat level area.
3. Membentuk fungsi *likelihood* untuk parameter model.
4. Menetapkan distribusi *prior* dari parameter model.
5. Membentuk *joint posterior distribution* berdasarkan fungsi *likelihood* dan distribusi *prior*.
6. *Joint Posterior Distribution* yang diperoleh digunakan untuk menentukan *full conditional posterior distribution* untuk masing-masing parameter.
7. Besaran *prior* dihitung menggunakan integrasi numerik dengan metode MCMC. Algoritma yang digunakan yaitu *Gibbs Sampling*. Penghitungan distribusi *posterior* akan menghasilkan sampel-sampel besaran *posterior*.

3.5. Aplikasi Model

Langkah-langkah yang akan dilakukan untuk tujuan kedua yakni mengaplikasikan metode HB dengan *Multinomial Logit Mixed Model* pada kasus lama sekolah di Kabupaten Keerom adalah sebagai berikut:

1. Studi literatur.
2. Melakukan *pre-processing* data dengan eksplorasi data dan membuat statistik deskriptif untuk variabel respon dan prediktor berdasarkan sumber data pada seluruh kecamatan di Kabupaten Keerom.
3. Melakukan estimasi langsung proporsi lama sekolah menurut kategori di setiap kecamatan di Kabupaten Keerom. Estimasi ini dilakukan sebagai pembanding nilai estimasi yang diperoleh dari hasil estimasi menggunakan model HB.
4. Melakukan estimasi parameter menggunakan metode MCMC dengan mengaplikasikan *Gibbs Sampling* serta menggunakan *joint posterior distribution* dan *full conditional posterior distribution* yang terbentuk.

Adapun langkah-langkahnya adalah:

- a. Menentukan nilai awal (*intial value*) untuk setiap parameter yang akan diestimasi.
 - b. Menentukan banyaknya iterasi (M) untuk membangkitkan sampel setiap parameter. Semakin besar jumlah iterasi yang dilakukan, maka hasil estimasinya akan semakin konvergen.
 - c. Membangkitkan sampel parameter sampai kondisi *equilibrium* tercapai. Jika belum tercapai sampai iterasi ke M, lakukan kembali proses iterasi hingga kondisi *equilibrium* tercapai.
 - d. Menghitung karakteristik distribusi *posterior* (mean, median, standar deviasi).
5. Menghitung koefisien variasi (CV) dari estimasi langsung maupun estimasi menggunakan model HB, lalu membandingkan hasilnya.
 6. Membandingkan hasil estimasi langsung dan estimasi menggunakan model HB.

Untuk menentukan apakah rantai Markov telah mencapai konvergensi atau belum, dapat dilakukan dengan pemeriksaan secara visual. Beberapa cara pendeteksian sebagai berikut (Lunn, 2013):

1. *History/Trace Plot*

Trace Plot merupakan plot dari iterasi versus nilai yang telah dibangkitkan. Ketika algoritma MCMC dimulai dengan nilai-nilai parameter yang jauh dari pusat distribusi target, maka pada awal iterasi algoritma akan menyimpang dari daerah ruang parameter dimana distribusi target dipusatkan. Fluktuasi dari nilai parameter pada *trace plot* menunjukkan bahwa *burn-in* belum tercapai. Jika fluktuasi ini muncul maka perlu untuk menghilangkan bagian awal dari rantai, karena nilai-nilai awal ini tidak menunjukkan perkiraan sampel yang benar dari distribusi target. Jadi bisa dikatakan bahwa jika semua nilai-nilai berada dalam daerah tanpa keperiodikan yang kuat cenderung dapat diasumsikan telah konvergen.

2. Plot Autokorelasi

Pada algoritma *Gibbs sampling*, nilai simulasi pada iterasi bergantung pada iterasi tepat sebelumnya. Jika pada rantai terdapat korelasi yang kuat di antara nilai-nilai yang berurutan maka kedua nilai berurutan tersebut memberikan informasi hanya secara marginal mengenai distribusi target dan bukan nilai dari sebuah simulasi tunggal. Korelasi yang kuat juga menunjukkan bahwa algoritma masih berada pada daerah tertentu dari ruang parameter dan mungkin membutuhkan waktu yang lama untuk penyampelan dari keseluruhan distribusi.

3. *Ergodic Mean Plot*

Merupakan plot antara iterasi dengan nilai *mean*-nya. Jika *ergodic mean* stabil, maka ini menjadi indikasi algoritma telah mencapai kekonvergenan.

(Halaman ini sengaja dikosongkan)

BAB IV

HASIL DAN PEMBAHASAN

4.1 Estimasi Parameter *Multinomial Logit Mixed Model* dengan pendekatan *Hierarchical Bayes*

Jika setiap hasil pengukuran dikategorikan dalam k kategori A_1, A_2, \dots, A_k , maka proporsi pada kategori ke- j , $j = 1, 2, \dots, k$ dinyatakan oleh p_j dan dinyatakan bahwa p_1, p_2, \dots, p_k adalah parameter dari sebaran multinomial. Untuk area tertentu (μ_i) tertentu sebaran peluang multinomial dapat dinyatakan sebagai:

$$y_{ij} | \mu_{ij} \sim \text{Multinomial}(n_i, p_{i1}, p_{i2}, \dots, p_{i(k-1)}) \quad (4.1)$$

dengan $i = 1, 2, \dots, m$ adalah banyaknya area, $j = 1, 2, \dots, k$ adalah banyaknya kategori, p_{ij} adalah probabilitas lama sekolah area ke- i pada kategori ke- j dengan $\sum_{j=1}^k p_{ij} = 1$, dan n_{ij} adalah jumlah sampel pada area ke- i pada kategori ke- j dengan $\sum_{j=1}^k n_{ij} = n_i$.

Misal \mathbf{z}_{ij} merupakan vektor penyerta, maka model linier yang didasarkan pada rasio $\theta_{ij} = \log \frac{p_{ij}}{p_i} = \log \frac{p_{ij}}{1 - \sum_{j=1}^k p_{ij}}$ yaitu:

$$\theta_{ij} = \mathbf{z}_{ij}^T \boldsymbol{\beta}_j + v_j, \quad (4.2)$$

untuk $j = 1, 2, \dots, k$, dan $i = 1, 2, \dots, m$, dengan v_i merupakan pengaruh acak area dan diasumsikan mengikuti distribusi multivariat normal dan saling bebas. Peluang dari kategori ke- j dalam area ke- i dapat dirumuskan:

$$p_{ij} = \frac{e(\theta_{ij})}{1 + \sum_{j=1}^{k-1} e(\theta_{ij})} \quad (4.3)$$

Membentuk kerangka HB model SAE untuk link model seperti pada persamaan (2.3) dengan model berbasis area hingga diperoleh model logit normal dengan kovariat pada level area. Pendekatan HB menggunakan distribusi marjinal dari setiap komponen Multinomial yaitu distribusi Binomial. Pendekatan HB untuk model seperti persamaan (4.2) dapat dituliskan seperti persamaan (4.4).

- (i) $y_{ij} | p_{ij} \sim \text{Binomial}(n_i, p_{ij})$
- (ii) $\theta_{ij} = \text{logit}(p_{ij}) = \mathbf{z}_{ij}^T \boldsymbol{\beta}_j + v_{ij}$, dengan $v_{ij} \sim N(0, \sigma_i^2)$
- (iii) $\boldsymbol{\beta}_j$ dan σ_i^2 saling bebas.

Penggunaan fungsi hubung *logit* (p_{ij}) sesuai dengan teori *Generalized Linear Model* (GLM) yang menyatakan bahwa fungsi hubung *logit* menjadi salah satu *link function* yang bisa digunakan untuk transformasi variabel respon sehingga hubungan antara variabel respon (proporsi) dengan variabel penyerta menjadi lebih tepat. Fungsi *logit* (p_{ij}) akan menjamin bahwa estimasi dari proporsi (p_{ij}) akan selalu jatuh pada daerah 0 dan 1.

Berdasarkan poin (ii) pada persamaan (4.4), jika $\theta_{ij} \sim N(\mathbf{z}_{ij}^T \boldsymbol{\beta}_j, \sigma_i^2)$, maka fungsi *likelihood* $L(\boldsymbol{\theta} | \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ dapat dijelaskan sebagai berikut:

$$f(\boldsymbol{\theta} | \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{j=1}^k (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\}, \quad (4.5)$$

maka fungsi *likelihood* dari $\boldsymbol{\theta}$ dinyatakan sebagai berikut:

$$\begin{aligned} L(\boldsymbol{\theta} | \boldsymbol{\beta}, \boldsymbol{\sigma}^2) &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{j=1}^k (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\} \\ &= \frac{1}{(2\pi\sigma_i^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{i=1}^m \sum_{j=1}^k (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\} \\ &\propto (2\sigma_i^{-2})^{m/2} \exp \left\{ -\frac{2\sigma_i^{-2}}{2} \sum_{i=1}^m \sum_{j=1}^k (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\} \end{aligned} \quad (4.6)$$

Tahapan selanjutnya yakni menentukan distribusi *prior* untuk parameter yang akan diestimasi. Dalam penelitian ini distribusi *prior* diperlukan untuk parameter $\boldsymbol{\beta}_j$ dan σ_i^2 . Pada penelitian ini digunakan distribusi *prior* yang bersifat independen, yaitu distribusi *prior* yang satu dengan lainnya saling bebas.

Distribusi *prior* ini memegang peranan penting karena akan mempengaruhi inferensi pada distribusi *prior*. Distribusi *prior* merupakan perpaduan antara distribusi *prior* dan data observasi yang digunakan untuk menyusun fungsi

likelihood. Proses untuk mendapatkan distribusi *prior* memerlukan proses analitik maupun integral numerik yang rumit dan sulit dipecahkan, namun hal ini dapat diatasi dengan menggunakan *Markov Chain Monte Carlo* (MCMC). Metode MCMC dilakukan dengan membangkitkan data parameter sesuai proses Markov Chain dengan simulasi Monte Carlo secara iteratif sehingga diperoleh distribusi *prior* yang stasioner. Kestasioneran distribusi *prior* ditandai dengan *Markov Chain* yang dihasilkan sudah konvergen. Gibbs sampling merupakan salah satu algoritma yang paling sederhana dari MCMC. Penarikan sampel dari distribusi *prior* melalui marginal distribusi sulit dilakukan sehingga dilakukan dengan pendekatan *full conditional*. *Full conditional distribution* didapatkan dengan cara mengeliminasi komponen dari distribusi *prior* gabungan yang bukan parameter target dengan menganggap parameter yang lain *fixed*.

Pada penelitian ini akan digunakan *flat prior* untuk β_j , yakni $f(\beta_j) \propto 1$. *Flat prior* adalah *prior* yang mempunyai densitas serba sama (*equally likely/uniformly distributed*). Jenis *prior* ini digunakan karena β_j hanya mempunyai nilai-nilai pada *range* tertentu saja. Interpretasinya adalah bahwa setiap kondisi diberi kesempatan yang sama untuk dapat terpilih sebagai *suporter likelihood* dalam membentuk *posterior*-nya (Iriawan, 2006). Sedangkan untuk parameter σ_i^2 akan digunakan distribusi *prior* konjugat, yaitu $\sigma_i^{-2} \sim \text{Gamma}(a, b)$ dengan $a \geq 0, b \geq 0$, dengan fungsi pdf:

$$\begin{aligned} f(\sigma_i^{-2}) &= \frac{b^a}{\Gamma(a)} (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2}b) \\ &\propto (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2}b) \end{aligned} \quad (4.7)$$

Bentuk distribusi *prior* yang bersifat independen dalam pemodelan HB dapat dituliskan sebagai berikut:

$$\begin{aligned} f(\beta, \sigma_{ij}^{-2}) &\propto f(\beta) * f(\sigma_i^{-2}) \\ f(\beta, \sigma_i^{-2}) &\propto 1 * (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2}b) \\ f(\beta, \sigma_i^{-2}) &\propto (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2}b) \end{aligned} \quad (4.8)$$

Pembentukan *joint posterior distribution* dari seluruh parameter yang akan diestimasi dilakukan dengan mengkombinasikan antara *likelihood* dan *prior*. Dari *likelihood* pada persamaan (4.6) dan distribusi *prior* pada persamaan (4.8) diperoleh *joint posterior distribution* seperti pada persamaan (4.9):

$$f(\boldsymbol{\beta}, \sigma_i^{-2} | \boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta} | \boldsymbol{\beta}, \sigma_i^{-2}) * f(\boldsymbol{\beta}, \sigma_i^{-2})}{h(\boldsymbol{\theta})}, \quad (4.9)$$

dimana $h(\boldsymbol{\theta}) = \int \dots \int L(\boldsymbol{\theta} | \boldsymbol{\beta}, \sigma_i^{-2}) * f(\boldsymbol{\beta}, \sigma_i^{-2}) \partial \beta_1 \dots \partial \beta_m \partial \sigma_{i1}^{-2} \dots \partial \sigma_i^{-2}$ dan merupakan konstanta densitas yang tidak tergantung pada parameter. Dengan demikian, maka persamaan (4.9) dapat dinyatakan dalam bentuk proporsional sebagai berikut:

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma_i^{-2} | \boldsymbol{\theta}) &\propto \frac{L(\boldsymbol{\theta} | \boldsymbol{\beta}, \sigma_i^{-2}) * f(\boldsymbol{\beta}, \sigma_i^{-2})}{h(\boldsymbol{\theta})} \\ f(\boldsymbol{\beta}, \sigma_i^{-2} | \boldsymbol{\theta}) &\propto (\sigma_i^{-2})^{m/2} \exp \left\{ -\frac{\sigma_i^2}{2} \sum_{i=1}^m \sum_{j=1}^k (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\} \\ &\quad * (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2} b). \end{aligned} \quad (4.10)$$

Selanjutnya ingin diperoleh distribusi *prior* marginal untuk setiap target parameter $\boldsymbol{\beta}$ dan σ_i^{-2} :

(i) Distribusi *prior* marginal parameter $\boldsymbol{\beta}$

$$\begin{aligned} f(\boldsymbol{\beta} | \sigma_i^{-2}, \boldsymbol{\theta}) &\propto (\sigma_i^{-2})^{m/2} \exp \left\{ -\frac{\sigma_i^2}{2} \sum_{i=1}^m \sum_{j=1}^k (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2 \right\} \\ f(\boldsymbol{\beta} | \sigma_i^{-2}, \boldsymbol{\theta}) &\propto L(\boldsymbol{\theta} | \boldsymbol{\beta}, \sigma_i^{-2}) \sim N(\mathbf{z}_{ij}^T \boldsymbol{\beta}_j, \sigma_i^{-2}), \text{ maka} \\ (\boldsymbol{\beta} | \sigma_i^{-2}, \boldsymbol{\theta}) &\sim N_p \left(\boldsymbol{\beta}^*, \sigma_v^2 \left(\sum_{i=1}^m \sum_{j=1}^k \mathbf{z}_{ij} \mathbf{z}_{ij}^T \right)^{-1} \right), \end{aligned} \quad (4.11)$$

$$\text{dengan } \boldsymbol{\beta}^* = \left(\sum_{i=1}^m \sum_{j=1}^k \mathbf{z}_{ij}^T \mathbf{z}_{ij} \right)^{-1} \left(\sum_{i=1}^m \sum_{j=1}^k \mathbf{z}_{ij}^T \boldsymbol{\theta}_{ij} \right).$$

Bentuk *full conditional posterior* untuk $\boldsymbol{\beta}_{ij}$ adalah

$$f(\boldsymbol{\beta}_{ij} | \sigma_i^{-2}, \boldsymbol{\theta}, \boldsymbol{\beta}_{\setminus ij}) = \int \dots \int f(\boldsymbol{\beta}, \sigma_i^{-2} | \boldsymbol{\theta}) \partial \beta_1 \dots \partial \beta_{ij-1} \partial \sigma_{i1}^{-2} \dots \partial \sigma_i^{-2},$$

dimana $\boldsymbol{\beta}_{\setminus ij}$ merupakan vektor tanpa elemen ke- j .

(ii) Distribusi *prior* marginal parameter σ_i^{-2}

$$f(\sigma_i^{-2} | \boldsymbol{\beta}, \boldsymbol{\theta}) \propto (\sigma_i^{-2})^{m/2} \exp\left\{-\frac{\sigma_i^{-2}}{2} \sum_{i=1}^m \sum_{j=1}^k (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right\} * (\sigma_i^{-2})^{a-1} \exp(-\sigma_i^{-2} b)$$

$$f(\sigma_i^{-2} | \boldsymbol{\beta}, \boldsymbol{\theta}) \propto (\sigma_i^{-2})^{\frac{m}{2}+a-1} \exp\left\{-\sigma_i^{-2} \left(b + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^k (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right)\right\},$$

maka:

$$(\sigma_i^{-2} | \boldsymbol{\beta}, \boldsymbol{\theta}) \sim \text{Gamma}\left(\frac{m}{2} + a, b + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^k (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right) \quad (4.12)$$

Bentuk *full conditional posterior* untuk σ_i^{-2} :

$$f(\sigma_i^{-2} | \boldsymbol{\theta}, \boldsymbol{\beta}_{ij}, \sigma_{i \setminus j}^{-2}) = \int \dots \int f(\boldsymbol{\beta}, \sigma_i^{-2} | \boldsymbol{\theta}) \partial \beta_1 \dots \partial \beta_j \dots \partial \sigma_{i1}^{-2} \dots \partial \sigma_{i-1}^{-2},$$

dimana $\sigma_{i \setminus j}^{-2}$ merupakan varians σ_j^{-2} tanpa elemen ke- j .

Terlihat dari distribusi *prior* marginal yang dihasilkan pada persamaan (4.11) dan (4.12) berbentuk *close form*, sehingga besaran *prior* dapat dihitung secara langsung menggunakan distribusi marginalnya menggunakan proses MCMC dengan algoritma *Gibbs Sampling*. Penghitungan distribusi *prior* akan menghasilkan sampel-sampel besaran *prior*. Dari ringkasan distribusi dapat diperoleh besaran mean, median, dan standar deviasi.

Diketahui fungsi densitas dari $f(p_{ij} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^{-2})$ dapat dirumuskan:

$$f(p_{ij} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^{-2}) \propto h(p_{ij} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^{-2}) * k(p_{ij}) \quad (4.13)$$

dimana:

$$h(p_{ij} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^{-2}) \propto g'(p_{ij}) \exp\left\{-\frac{1}{2\sigma_i^2} (\boldsymbol{\theta}_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right\}, \text{ dan } k(p_{ij}) = p_{ij}^y (1 - p_{ij}^y)^{n_i - y_i}$$

dengan $g'(p_{ij}) = \partial g(p_{ij}) / \partial p_{ij}$ dan $g(p_{ij}) = \text{logit}(p_{ij})$.

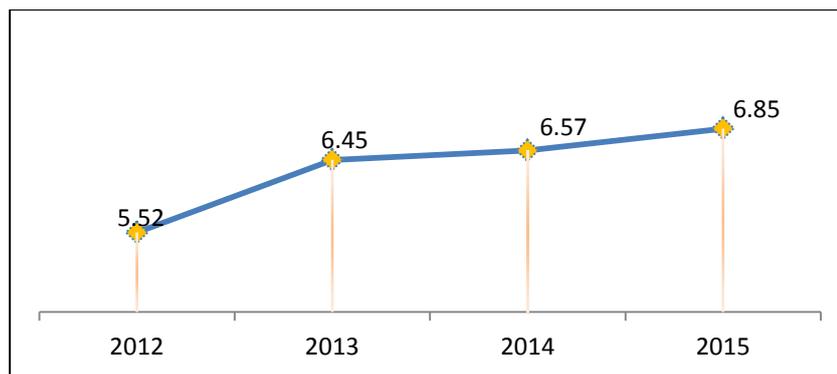
Dari persamaan (4.13) diketahui bahwa distribusi dari $(p_{ij} | \boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_i^{-2})$ bentuknya tidak *close form* sehingga untuk memperoleh estimasi dari p_{ij} dilakukan dengan proses MCMC. Estimasi model HB untuk p_{ij} dan varians dari *prior* p_{ij}

dihasilkan secara langsung dari sampel MCMC $\{p_1^{(j)}, \dots, p_m^{(j)}, \beta_j^{(j)}, \sigma_i^{2(j)}; j = d + 1, \dots, d + D\}$ yang diperoleh dari *joint prior* $f(p_1, \dots, p_m, \beta_j, \sigma_i^2 | y)$, dimana d banyaknya proses *burn-in* dan D adalah banyaknya sampel yang diharapkan.

4.2 Gambaran Umum Kabupaten Keerom

Kabupaten Keerom terletak antara 2 37' 0" – 4 0' 0" LS dan antara 140 15' 0" – 141 0' 0" BT, dengan luas wilayah sebesar 9.365 km², dan terbagi menjadi 7 kecamatan dengan 4 kecamatan diantaranya berbatasan langsung dengan Negara Papua New Guinea.

Dari hasil proyeksi penduduk tahun 2015, terdapat sebanyak 13.026 rumah-tangga, jumlah penduduk mencapai 53.694 jiwa, dengan rata-rata jumlah anggota rumahtangga sebesar 4,12. Persebaran penduduk di wilayah ini relatif kurang merata, dengan rata-rata kepadatan penduduk sebesar 6 Jiwa/Km², terdapat beberapa daerah seperti Kecamatan Senggi dan Web yang memiliki kepadatan penduduk berada di bawah rata-rata yaitu masing-masing sebesar 1 Jiwa/Km² dan 3 Jiwa/Km². Sedangkan beberapa kecamatan tergolong sangat padat, seperti di Kecamatan Arso, yang merupakan ibukota kabupaten, dan Kecamatan Arso Timur, dengan kepadatan penduduknya masing-masing sebesar 13 Jiwa/Km² dan 11 Jiwa/Km².



Gambar 4.1. Rata-rata Lama Sekolah Kabupaten Keerom Tahun 2012 – 2015

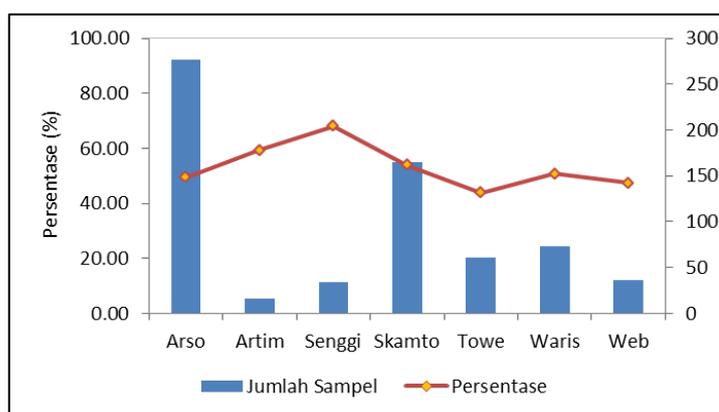
Hasil penghitungan BPS menunjukkan bahwa RLS Kabupaten Keerom tahun 2015 sebesar 6,85. Hal ini menunjukkan bahwa rata-rata penduduk Kabupaten Keerom bersekolah tidak sampai menyelesaikan kelas 1 SMP/Sederajat. Angka ini meningkat jika dibandingkan tahun 2014, yaitu sebesar 6,57. Walaupun ada

kenaikan signifikan dari RLS Kabupaten Keerom, tetapi program pendidikan wajib belajar 9 tahun belum berhasil.

Sebelum dilakukan pemodelan, terlebih dahulu dilakukan verifikasi variabel yang akan dimasukkan dalam model. Dari data Susenas 2015 Kor Individu sebanyak 662 individu dilakukan penghitungan antar variabel untuk mendapatkan variabel respon. Adapun untuk variabel prediktor bersumber dari Podes 2014 dan publikasi yang diterbitkan oleh BPS Kabupaten Keerom. Banyaknya pengamatan adalah sebanyak 7 kecamatan.

Berdasarkan hasil Susenas Tahun 2015 di Kabupaten Keerom diperoleh informasi bahwa total sampel sebanyak 1.302 jiwa, dengan penduduk usia 25 tahun ke atas sebanyak 662 orang, dengan persentase jumlah sampel terhadap total sampel Susenas di Kabupaten Keerom sebesar 50,84 persen.

Jumlah sampel penduduk usia 25 tahun ke atas tiap kecamatan berbeda. Hal ini disebabkan sampel Susenas tidak dikhususkan untuk memperoleh informasi mengenai pendidikan. Susenas bertujuan untuk memperoleh informasi sosial kependudukan yang relatif sangat luas terutama terkait dengan tingkat kesejahteraan, dimana di dalamnya terdapat pendidikan. Hal tersebut menjadi salah satu alasan adanya perbedaan jumlah sampel dan persentasenya di setiap kecamatan.



Gambar 4.2. Jumlah Sampel dan Persentase Sampel Penduduk Usia 25 tahun ke atas terhadap Sampel Susenas

Dari Gambar 4.2 terlihat bahwa Kecamatan Arso memiliki jumlah sampel Susenas yang teridentifikasi penduduk usia 25 tahun ke atas lebih besar dibandingkan kecamatan lainnya, namun jika dibandingkan dengan jumlah sampel Susenas masih memiliki persentase yang cukup kecil (49,46 persen). Persentase

jumlah sampel Susenas yang teridentifikasi penduduk usia 25 tahun ke atas lebih besar dibandingkan kecamatan lain terdapat di Kecamatan Senggi dan Arso Timur, masing-masing sebesar 68 persen dan 59,26 persen, dengan jumlah sampel yang justru lebih sedikit dibandingkan kecamatan lain.

Salah satu cara untuk mensukseskan program wajib belajar 9 tahun yaitu ketersediaan sarana dan prasarana pendidikan yang memadai, antara lain ketersediaan sekolah. Pada tahun 2015, tercatat sebanyak 8.390 siswa SD/Sederajat, 2.722 siswa SMP/Sederajat, dan 1.595 siswa SMA/Sederajat. Banyaknya jumlah murid tersebut tidak dibarengi dengan ketersediaan sekolah di tingkat kecamatan maupun desa. Minimnya ketersediaan sekolah di tingkat kecamatan maupun desa diakibatkan letak geografis empat dari tujuh kecamatan di Kabupaten Keerom yang sulit dijangkau, yaitu Kecamatan Senggi, Web, Waris dan Towe. Topografi, jumlah sekolah dan jarak ibukota kecamatan ke ibukota kabupaten di Kabupaten Keerom pada tahun 2015 dapat dilihat pada Tabel 4.1.

Tabel 4.1. Topografi, Jumlah Sekolah dan Jarak berdasarkan Kecamatan di Kabupaten Keerom Tahun 2015

Kecamatan	Topografi (X1)	Jumlah SD (X2)	Jumlah SMP (X3)	Jumlah SMA (X4)	Jarak (km) (X5)
Arso	1	23	5	7	2,4
Arso Timur	1	11	3	0	27,6
Senggi	1	10	1	1	95,1
Skamto	1	15	4	4	32,4
Waris	2	5	1	1	74,4
Web	2	6	1	0	139,8
Towe	2	6	0	0	185,4
Jumlah		76	17	13	

Sumber : Badan Pusat Statistik Kabupaten Keerom (2016)

Dari Tabel 4.2 diperoleh informasi bahwa tidak semua kategori lama sekolah di setiap kecamatan memiliki sampel pada Susenas, yang akan mengakibatkan informasi untuk kategori tersebut tidak tersedia jika dilakukan estimasi langsung. SAE dapat dimanfaatkan untuk jumlah sampel yang terbatas.

Tabel 4.2. Jumlah Sampel Penduduk Usia 25 tahun ke atas menurut Kategori Lama Sekolah

Kecamatan	Kategori						Jumlah
	1	2	3	4	5	6	
Arso	48	27	45	53	80	24	277
Arso Timur	0	0	8	4	3	1	16
Senggi	6	0	6	11	9	2	2
Skamto	28	16	30	45	35	11	11
Waris	33	8	16	6	8	2	73
Web	14	8	3	7	4	0	36
Towe	56	1	1	1	0	2	61
Keerom	185	60	109	127	141	40	662

Sumber : BPS Kabupaten Keerom (2016)

Sebelum melakukan estimasi dengan model HB, dilakukan estimasi secara langsung terhadap proporsi lama sekolah menurut kategori di setiap kecamatan. Nilai yang diperoleh dalam estimasi akan digunakan sebagai pembandingan dengan nilai estimasi yang diperoleh dari hasil HB. Hasil estimasi langsung proporsi lama sekolah menurut kategori disajikan pada Tabel 4.3. Jumlah kecamatan di Kabupaten Keerom pada tahun 2015 tercatat sebanyak 7 kecamatan.

Tabel 4.3. Estimasi Langsung Proporsi Lama Sekolah berdasarkan Kecamatan di Kabupaten Keerom, Tahun 2015.

Kecamatan	Kategori					
	1	2	3	4	5	6
1. Arso	0,1588	0,0830	0,1625	0,2202	0,2780	0,0975
2. Arso Timur	0,0000	0,0000	0,5000	0,2500	0,1875	0,0625
3. Senggi	0,1765	0,0000	0,1765	0,3235	0,2647	0,0588
4. Skamto	0,1515	0,0727	0,2061	0,2909	0,2061	0,0727
5. Waris	0,4521	0,0959	0,2329	0,0822	0,1096	0,0274
6. Web	0,3333	0,1667	0,1944	0,1944	0,1111	0,0000
7. Towe	0,9180	0,0164	0,0164	0,0164	0,0328	0,0000

Sumber : Hasil Olah

Pada tabel 4.3 dapat dilihat bahwa estimasi langsung proporsi lama sekolah di Kecamatan Arso Timur untuk kategori 1 dan 2, Kecamatan Senggi untuk kategori 2, dan pada kategori 6 pada kecamatan Towe dan Web adalah 0. Hal ini dikarenakan keterbatasan jumlah sampel pada kecamatan tersebut di kategori tertentu.

Hal ini yang menjadi salah satu dasar dalam menggunakan metode SAE dengan pendekatan HB.

4.3 Estimasi Proporsi Lama Sekolah menggunakan Metode *Hierarchical Bayes (HB)* dengan *Multinomial Logit Mixed Model*

Informasi mengenai proporsi lama sekolah belum tersedia secara lengkap di setiap kecamatan, sehingga diperlukan metode estimasi lain untuk melengkapi informasi lama sekolah di seluruh kecamatan. Metode SAE dengan pendekatan HB menggunakan *Multinomial Logit Mixed Model* akan digunakan untuk mengestimasi proporsi lama sekolah menurut kategori lama sekolah di setiap kecamatan di Kabupaten Keerom.

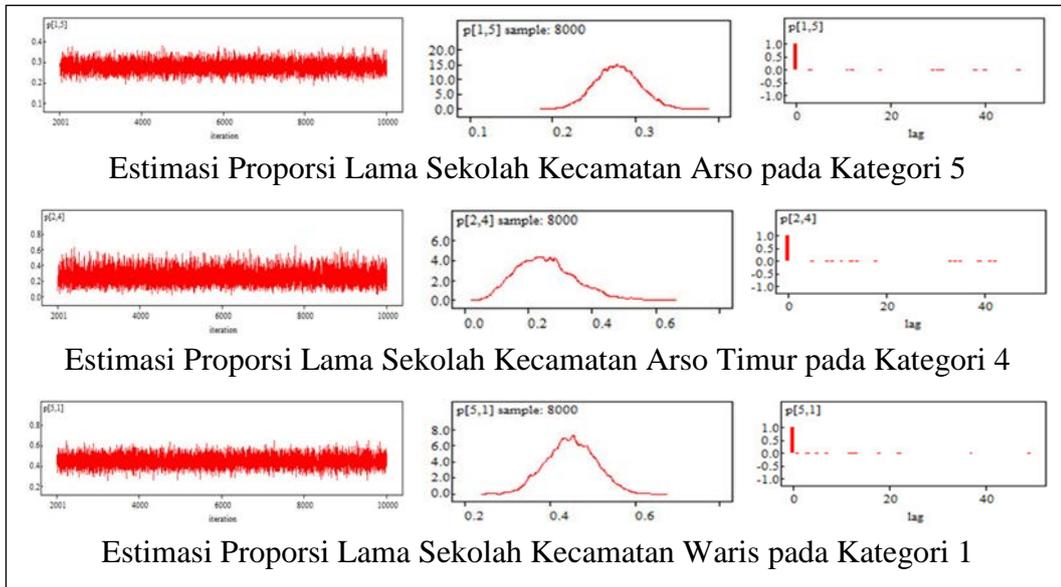
Dalam melakukan estimasi p_i , terlebih dahulu dilakukan estimasi terhadap β dan σ_i^2 melalui metode MCMC dengan algoritma *Gibbs Sampling*. *Prior* yang digunakan pada penelitian ini menggunakan *pseudo prior* untuk β_j , yakni *prior* dengan pemberian nilainya disetarakan dengan hasil elaborasi dari frekuentis. Sedangkan untuk parameter σ_i^2 akan digunakan distribusi *prior* konjugat, yaitu $\sigma_i^2 \sim \text{Gamma}(a, b)$, dengan $a \geq 0, b \geq 0$. Dikarenakan tidak adanya informasi awal, maka parameter distribusi Gamma ditetapkan sebesar $a = 8$ dan $b = 8,1$.

Hal berikutnya yang dapat dilakukan adalah dengan memasukkan variabel respon dan variabel penyerta ke dalam model serta melakukan estimasi parameter model. Proses estimasi parameter model dilakukan melalui proses MCMC, yakni proses membangun suatu peluang rantai Markov hingga menuju distribusi *prior* tertentu. Dari distribusi *prior* tersebut dapat diperoleh karakteristik distribusi, sehingga diperoleh estimasi parameter model.

Proses MCMC dilakukan dengan cara iterasi. Setiap iterasi, masing-masing parameter akan menghasilkan nilai yang baru. Nilai estimasi parameter diperoleh dari rata-rata nilai setelah rantai Markov konvergen. Semakin kompleks suatu model, iterasi yang diperlukan semakin banyak yang mengakibatkan *lag* autokorelasi dalam sampel Markov chain yang dibangkitkan semakin panjang. Panjangnya *lag* autokorelasi merupakan identifikasi bahwa pergerakan nilai parameter tidak

mengikuti sifat Markov chain yang *irreducible*, *aperiodic*, dan *reccurent*. Perbesaran nilai *thin* diperlukan untuk memperoleh sampel Markov chain yang independen. Sebagai ilustrasi, jika *thin* = 20, maka hanya sampel urutan iterasi ke 20, 40, dan seterusnya yang akan menjadi sampel. Proses *burn-in* juga dilakukan jika pada nilai-nilai awal pada rantai Markov belum konvergen. Proses *burn-in* adalah proses menghilangkan bagian awal dari rantai Markov karena belum menunjukkan perkiraan sampel yang akan menuju pada distribusi tertentu. Nilai estimasi parameter diperoleh dari rata-rata setelah suatu rantai Markov konvergen.

Pada penelitian ini, kekonvergenan rantai Markov diperoleh setelah proses *burn-in* sebanyak 2.000 dari 10.000 iterasi yang dilakukan, dengan jumlah *thin* sebesar 5. Diagnosa konvergen pada rantai Markov dapat dilakukan dengan pemeriksaan *trace plot*, *density plot*, dan *plot* autokorelasi. Hasil *trace plot* menunjukkan rantai Markov telah konvergen ketika nilai estimasi parameter sudah tidak membentuk pola naik turun. Kekonvergenan rantai Markov dilihat dari *density plot* jika sudah menunjukkan pola distribusi yang mulus, dan dari *plot* autokorelasi yang sudah menunjukkan sampel yang dihasilkan pada Markov chain telah independen, atau tidak tergantung pada keadaan sebelumnya.

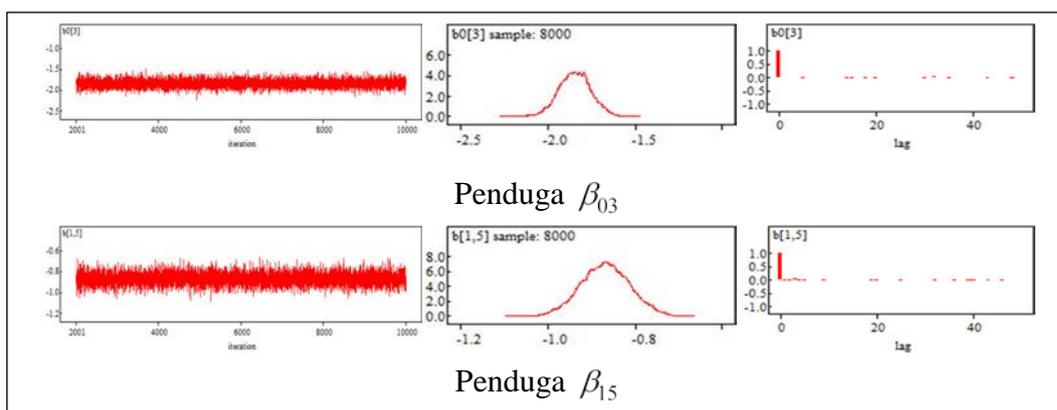


Gambar 4.3. *Plot* distribusi *Posterior* Parameter p_1^{HB} , p_2^{HB} , dan p_5^{HB}

Pada Gambar 4.3, kolom pertama menunjukkan *trace plot*, kolom kedua menunjukkan *density plot*, dan kolom ketiga menunjukkan *autocorrelation plot*

pada Kecamatan Arso, Arso Timur dan Waris. *Trace plot* untuk nilai proporsi lama sekolah pada ketiga kecamatan tersebut sudah tidak membentuk pola naik atau turun. Ini menunjukkan proses *burn-in* telah selesai. Pada *density plot* (pola kepekatan peluang) menunjukkan pola bahwa distribusi dari estimasi parameter yang diamati cenderung simetris. Sedangkan pada *autocorrelation plot* terlihat pada nilai-nilai autokorelasi menunjukkan cut off sejak lag 0 yang menandakan antar sampel MCMC sudah independen. Batas iterasi ke-10.000 merupakan nilai periode *burn-in* yang mengindikasikan kovergensi algoritma dari sampel yang dibangkitkan.

Selain estimasi proporsi lama sekolah per kecamatan, proses MCMC dari model HB juga menghasilkan *output* lain berupa parameter regresi (β) dan varians antar area kecil (σ_i^2). Hal ini dapat dilihat pada Lampiran 2. Dari Lampiran 2 dapat dilihat bahwa semua parameter yang akan digunakan untuk memprediksi proporsi lama sekolah sudah diestimasi dengan akurat, dilihat dari nilai *MC error* yang sangat kecil yaitu di bawah 0,1% (Ntzoufras, 2009). Terlihat dari nilai *credible interval* 95% bahwa semua variabel penyerta telah signifikan, artinya bahwa variabel penyerta memberikan informasi tambahan pada proses estimasi untuk proporsi lama sekolah menurut kategori.



Gambar 4.4. Taksiran Parameter β_{03} dan β_{15} estimasi proporsi lama sekolah

Pada Gambar 4.4, kolom pertama menunjukkan *trace plot*, kolom kedua menunjukkan *density plot*, dan kolom ketiga menunjukkan *autocorrelation plot*. Proses iterasi parameter β menunjukkan kondisi yang relatif stasioner sejak awal

iterasi hingga iterasi ke-10.000. Ketiga jenis plot menunjukkan kondisi yang serupa dengan kondisi plot pada pendugaan proporsi lama sekolah (Gambar 4.3). Oleh karena itu, pendugaan nilai β dapat diperoleh dari distribusi pada iterasi ke-2001 s.d. iterasi ke-10.0000. Estimasi proporsi lama sekolah per kecamatan menggunakan metode HB dapat dilihat pada Tabel 4.4.

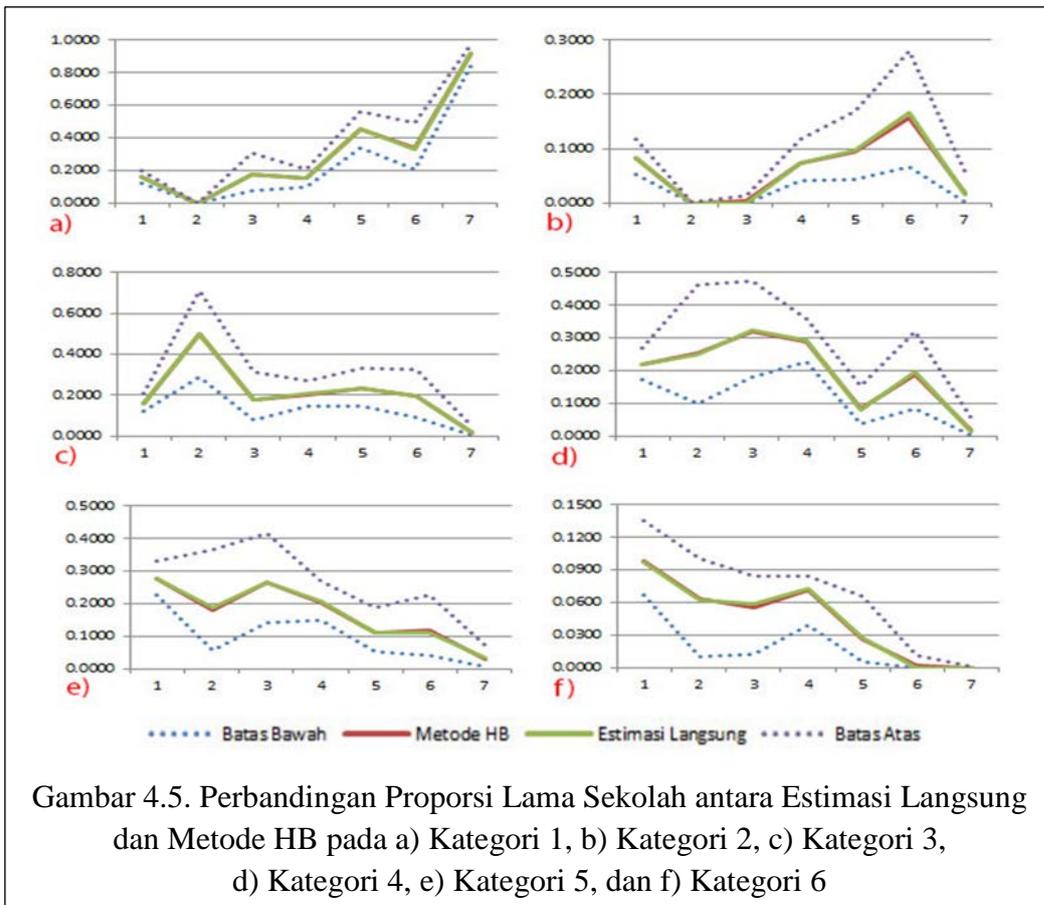
Tabel 4.4. Estimasi Proporsi Lama Sekolah menurut Kecamatan menggunakan Metode HB, Tahun 2015

Kecamatan	Kategori					
	1	2	3	4	5	6
1. Arso	0,1593	0,0823	0,1624	0,2207	0,2777	0,0977
2. Arso Timur	0,0002	0,0006	0,4996	0,2562	0,1789	0,0645
3. Senggi	0,1783	0,0031	0,1781	0,3193	0,2661	0,0552
4. Skamto	0,1516	0,0744	0,2057	0,2905	0,2062	0,0717
5. Waris	0,4486	0,0960	0,2340	0,0838	0,1110	0,0266
6. Web	0,3402	0,1602	0,1913	0,1872	0,1183	0,0027
7. Towe	0,9129	0,0193	0,0169	0,0204	0,0303	0,0002

Sumber : Hasil Olah

Tabel 4.4 menunjukkan hasil estimasi yang diperoleh dengan menggunakan model HB dengan *Multinomial Logit Mixed Model*. Jika dibandingkan dengan estimasi langsung terlihat bahwa terjadi perbedaan estimasi yang dihasilkan. Untuk sampel yang sedikit atau bahkan tidak ada, dengan menggunakan metode HB dapat diestimasi proporsi lama sekolah pada kecamatan tersebut.

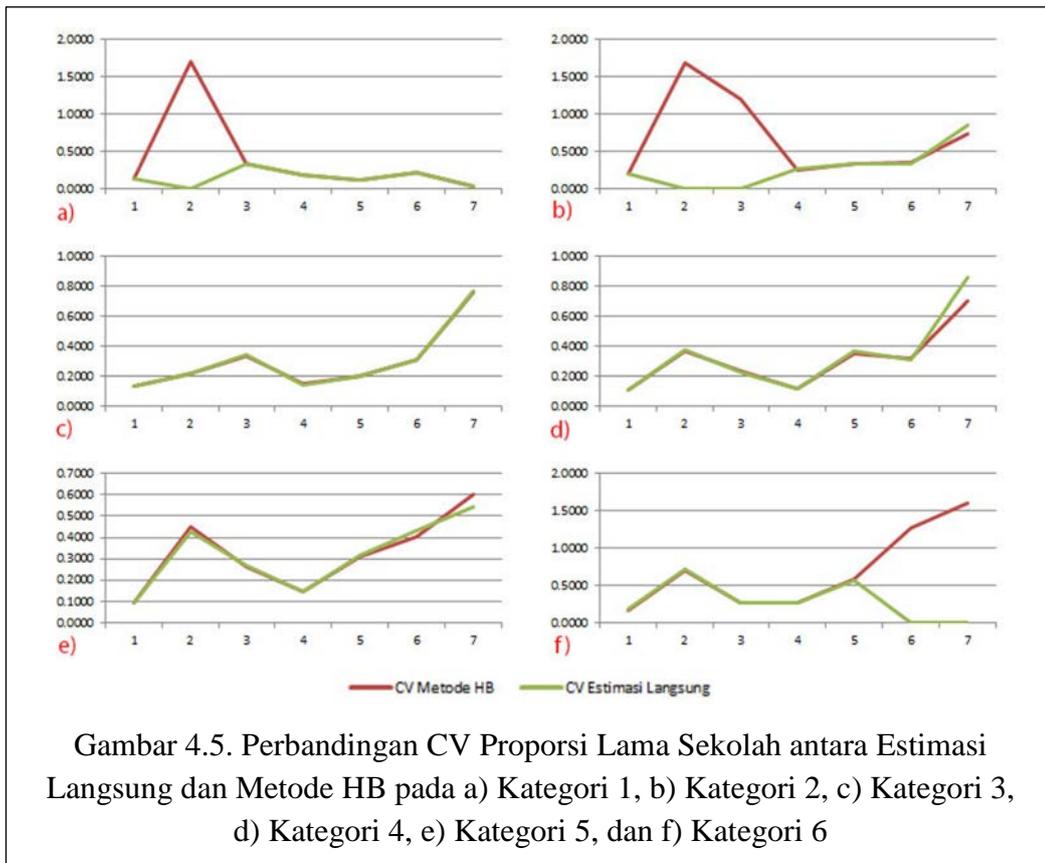
Nilai estimasi proporsi menggunakan HB menghasikan hasil yang bervariasi, beberapa diantaranya bernilai lebih tinggi jika dibandingkan nilai estimasi langsung dan beberapa lainnya lebih rendah. Perbedaan yang terjadi menunjukkan koreksi dari pendekatan metode HB terhadap nilai estimasi langsung. Koreksi dilakukan karena lemahnya akurasi estimasi langsung yang memiliki varians lebih besar sebagai akibat kurangnya sampel. Sebagai contoh, untuk proporsi pada kategori kedua, Kecamatan Arso Timur dan Senggi tidak memiliki sampel pada kategori ini sehingga estimasi langsung tidak dapat dilakukan. Dengan menggunakan metode HB, maka estimasi proporsi lama sekolah untuk kategori ini dapat dilakukan.



Gambar 4.5. Perbandingan Proporsi Lama Sekolah antara Estimasi Langsung dan Metode HB pada a) Kategori 1, b) Kategori 2, c) Kategori 3, d) Kategori 4, e) Kategori 5, dan f) Kategori 6

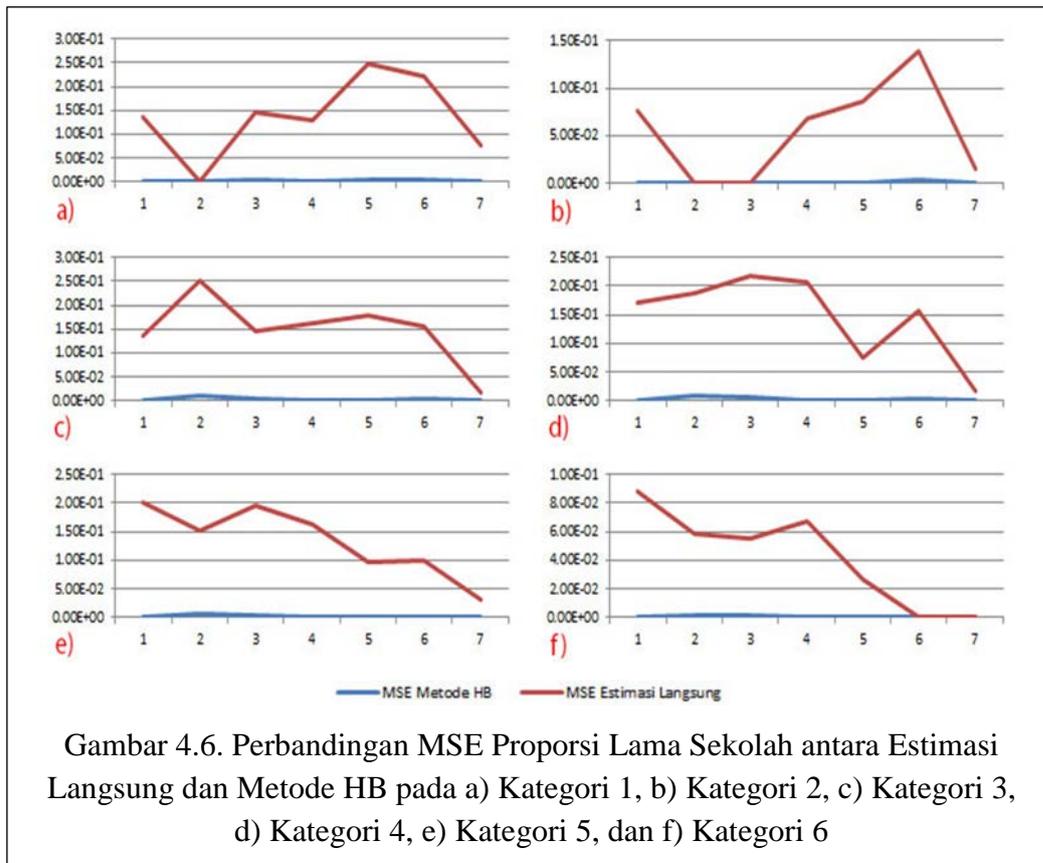
4.4 Perbandingan Koefisien Variasi (CV)

Umumnya nilai CV proporsi lama sekolah dengan metode HB lebih kecil dibandingkan estimasi langsung. Hal ini dapat dilihat misalnya pada Gambar 4.5 bagian d) dimana jelas terlihat nilai CV dengan metode HB lebih kecil dibandingkan nilai CV dengan estimasi langsung. Tetapi saat sampel kecil atau bahkan pada kondisi tidak ada sampel, nilai CV pada metode HB lebih besar dibandingkan estimasi langsung. Hal ini dapat dilihat pada Gambar 4.5 bagian a) pada Kecamatan Arso Timur, bagian b) pada Kecamatan Arso Timur dan Senggi, dan bagian f) pada Kecamatan Web dan Towe. Hal ini menunjukkan bahwa metode *hierarchical bayes multinomial logit normal* lebih baik dari hasil estimasi langsung dari sisi variansi dugaannya.



4.5 Perbandingan Mean Square Error (MSE)

Berdasarkan Gambar 4.6 terlihat bahwa dalam mengestimasi proporsi lama sekolah menurut kategori, MSE dari estimasi langsung cenderung lebih tinggi dibandingkan MSE menggunakan metode HB. Nilai MSE yang kecil pada estimasi menggunakan metode HB menggambarkan bahwa estimasi menggunakan metode HB menghasilkan kualitas estimasi yang lebih baik dalam mengestimasi proporsi lama sekolah menurut kategori. Adanya variabel penyerta dapat meningkatkan akurasi dalam mengestimasi proporsi lama sekolah. MSE metode HB menghasilkan nilai yang relatif sangat kecil, antara $9,58 \times 10^{-8}$ sampai $1,2 \times 10^{-2}$.



Gambar 4.6. Perbandingan MSE Proporsi Lama Sekolah antara Estimasi Langsung dan Metode HB pada a) Kategori 1, b) Kategori 2, c) Kategori 3, d) Kategori 4, e) Kategori 5, dan f) Kategori 6

4.6 Pembahasan

Estimasi model SAE pada respon multinomial dapat dilakukan menggunakan metode HB. Penggunaan metode HB dirasa sangat tepat karena datanya berbentuk binari atau cacahan. Penggunaan *link function* berupa fungsi logit sangat bermanfaat karena akan menjamin estimasi dari proporsi akan selalu berada pada daerah 0 dan 1. Penggunaan distribusi *prior* juga menjadi sangat penting dalam menentukan distribusi *prior*, karena akan mempengaruhi inferensi pada distribusi *prior*. Proses untuk mendapatkan inferensi pada distribusi *prior* memerlukan proses analitik maupun integral numerik yang rumit dan sulit dipecahkan, namun hal ini dapat diatasi dengan menggunakan *Markov Chain Monte Carlo* (MCMC). Pembentukan *joint posterior distribution* dari seluruh parameter yang akan diestimasi dilakukan dengan mengkombinasikan antara *likelihood* dan *prior*.

Pada kasus estimasi proporsi dengan respon multinomial dapat dituliskan sebagai berikut:

1. Fungsi *Likelihood*

$$L(\boldsymbol{\theta}|\boldsymbol{\beta}, \boldsymbol{\sigma}_v^2) \propto (\boldsymbol{\sigma}_v^2)^{m/2} \exp\left\{-\frac{\boldsymbol{\sigma}_v^2}{2} \sum_{i=1}^m \sum_{j=1}^q (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right\}$$

2. Distribusi *Prior*

$$f(\boldsymbol{\beta}|\boldsymbol{\sigma}_v^2, \boldsymbol{\theta}) \propto f(\boldsymbol{\beta}) * f(\boldsymbol{\sigma}_v^2) \propto (\boldsymbol{\sigma}_v^2)^{a-1} \exp(-\boldsymbol{\sigma}_v^2 b)$$

3. Distribusi *Prior Marginal*

$$(\boldsymbol{\beta}|\boldsymbol{\sigma}_v^2, \boldsymbol{\theta}) \sim N_p\left(\boldsymbol{\beta}^*, \boldsymbol{\sigma}_v^2 \left(\sum_{i=1}^m \sum_{j=1}^q \mathbf{z}_{ij}^T \mathbf{z}_{ij}\right)^{-1}\right),$$

$$\text{dengan } \boldsymbol{\beta}^* = \left(\sum_{i=1}^m \sum_{j=1}^q \mathbf{z}_{ij}^T \mathbf{z}_{ij}\right)^{-1} \left(\sum_{i=1}^m \sum_{j=1}^q \mathbf{z}_{ij}^T \boldsymbol{\theta}_{ij}\right),$$

$$(\boldsymbol{\sigma}_v^2|\boldsymbol{\beta}, \boldsymbol{\theta}) \sim \text{Gamma}\left(\frac{m}{2} + a, b + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^q (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right)$$

4. *Joint Distribution* untuk Estimasi Proporsi

$$f(p_{ij}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\sigma}_v^2) \propto h(p_{ij}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\sigma}_v^2) * k(p_{ij})$$

$$h(p_{ij}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\sigma}_v^2) \propto g'(p_{ij}) \exp\left\{-\frac{1}{2\boldsymbol{\sigma}_v^2} (\theta_{ij} - \mathbf{z}_{ij}^T \boldsymbol{\beta}_j)^2\right\}$$

$$k(p_{ij}) = p_{ij}^y (1 - p_{ij}^y)^{n_i - y_i}$$

$$\text{dengan } g'(p_{ij}) = \partial g(p_{ij}) / \partial p_{ij}, \text{ dan } g(p_{ij}) = \text{logit}(p_{ij}).$$

Distribusi *prior* yang telah berbentuk *close form* seperti pada parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\sigma}_v^{-2}$, maka besaran dari distribusi *prior*-nya dapat dihitung secara langsung menggunakan distribusi marginalnya menggunakan proses MCMC dengan algoritma *Gibbs Sampling*. Sedangkan untuk distribusi *prior* yang tidak berbentuk *close form*, seperti pada parameter p_{ij} , maka besaran *prior*-nya dihasilkan secara langsung dari sampel MCMC menggunakan *full conditional distribution*-nya. Estimasi pada setiap parameter diperoleh dari ringkasan statistik distribusi *prior*-nya, seperti mean, median, standar deviasi dan sebagainya.

Aplikasi dari model SAE menggunakan metode HB dilakukan untuk mengestimasi proporsi lama sekolah menurut kategori di setiap kecamatan di Kabupaten Keerom. Jumlah sampel yang teridentifikasi sebagai penduduk usia 25 tahun ke atas yang dihasilkan dari sampel Susenas di Kabupaten Keerom sangat bervariasi, dan beberapa kategori di kecamatan tidak memiliki sampel yang teridentifikasi masuk ke dalam kategori tersebut.

Nilai estimasi proporsi yang dihasilkan menggunakan HB bervariasi, beberapa diantaranya bernilai lebih tinggi jika dibandingkan nilai estimasi langsung dan beberapa lainnya lebih rendah (Gambar 4.5). Perbedaan yang terjadi menunjukkan koreksi dari pendekatan metode HB terhadap nilai estimasi langsung. Koreksi dilakukan karena lemahnya akurasi estimasi langsung yang memiliki varians lebih besar sebagai akibat kurangnya sampel. Adanya variabel penyerta memberikan pengaruh yang signifikan dan baik untuk mengestimasi proporsi lama sekolah menurut kategori.

BAB 5.

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan pada bab sebelumnya, maka kesimpulan yang dapat diperoleh sebagai berikut:

1. Variabel penyerta berpengaruh secara signifikan terhadap prediksi proporsi lama sekolah di Kabupaten Keerom.
2. Estimasi menggunakan metode HB mengkoreksi beberapa titik dari estimasi langsung. Koreksi terjadi karena lemahnya akurasi estimasi langsung yang memiliki varians lebih besar akibat kurangnya sampel.
3. Terdapat penurunan nilai CV dari estimasi menggunakan model HB dibandingkan estimasi langsung.
4. Dengan menggunakan metode Jackknife, nilai MSE pada model SAE menggunakan metode HB di setiap kategori lama sekolah menghasilkan nilai yang lebih kecil dibandingkan pada estimasi langsung. Hal tersebut mengindikasikan bahwa estimasi menggunakan metode HB dapat memperbaiki estimasi yang diperoleh menggunakan estimasi langsung.

5.2. Saran

Dalam penelitian ini masih banyak terdapat permasalahan yang belum dikaji secara mendalam, serta masih terdapat banyak keterbatasan di dalamnya, sehingga penulis memberikan saran sebagai berikut:

1. Pada penelitian ini banyak data yang bernilai 0, maka penelitian selanjutnya dapat menggunakan distribusi yang bersifat *truncated* sehingga data yang bernilai 0 dapat diestimasi secara tersendiri.
2. Pemilihan variabel penjelas yang tepat dan memiliki keterkaitan yang tinggi dengan variabel respon tentu akan memberikan akurasi yang tinggi untuk model *Hierarchical Bayes*. Pada penelitian selanjutnya disarankan untuk mengeksplorasi variabel penjelas lain yang lebih tepat untuk menjelaskan variabel lama sekolah.

3. Sesuai dengan kebutuhan informasi yang semakin diminati untuk wilayah kecil maupun domain yang lebih rinci, SAE dapat dimanfaatkan untuk memperoleh informasi yang lebih akurat dengan tidak memerlukan penambahan sampel yang tentunya memerlukan tambahan biaya yang lebih besar, waktu dan tenaga yang lebih banyak.

DAFTAR PUSTAKA

- Ayuningtyas, I. (2017), *Small Area Estimation* pada Kasus Respon Multinomial dengan Pendekatan *Hierarchical Bayes*. Aplikasi pada Proporsi Pengangguran menurut Kategori Pengangguran di Pulau Kalimantan, 2017, Tesis. Institut Teknologi Sepuluh Nopember, Surabaya.
- Baillo, A. dan Molina, I. (2009), *Mean-Squared Errors of Small Area Estimators under a Unit-Level Multivariate Model*, *Statistics*, Vol. 43, No. 6, pp. 553-569.
- Benavent, R., dan Morales, D. (2016). *Multivariate Fay-Herriot models for Small Area Estimation*. *Computational Statistics and Data Analysis*. Volume 94. p.372-390.
- Box, G.E. dan Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company, Inc.
- BPS Kabupaten Keerom (2016), Indeks Pembangunan Manusia Tahun 2015, Arso.
- BPS (2015), Survei Sosial Ekonomi Nasional (Susenas Maret 2015): Pedoman Pencacahan, BPS, Jakarta.
- Carlin, B. dan Chib, S. (1995), *Bayesian Model Choice via Markov Chain Monte Carlo Methods*, *Journal of the Royal Statistical Society: Series B Statistical Methodology*, Vol. 57(3), pp. 473-484.
- Casella, G. dan George, E.I. (1992), *Explaining the Gibbs Sampler*, *The American Statistician*, Vol. 46, No. 3, pp. 167-174.
- Chandra, H., Chambers, R. dan Salvati, N. (2009), *Small Area Estimation of Proportions in Business Surveys*, The University of Wollongong, Australia.
- Cochran, W. (1977), *Sampling Techniques*, Canada: John Wiley & Sons, Inc.
- Fausi, H. dan Sutikno (2011), *Small Area Estimation* terhadap Pengeluaran per Kapita di Kabupaten Sumenep dengan Metode *Empirical Bayes*, Skripsi: Institut Teknologi Sepuluh Nopember, Surabaya.
- Fay, R.E. dan Herriot, R.A. (1979), *Estimation of Income for Small Places : An Application of James-Stein Procedures to Census Data*, *Journal of the American Statistical Association*, pp. 743(366a), 266-277.
- Gelman, A., dkk. (2014). *Bayesian Data Analysis, 3rd Edition*. Boca Raton, FL, USA: Chapman & Hall/CRC.
- Ghosh, M. dan Rao, J.N.K. (1994), *Small Area Estimation: An Appraisal*. *Statistical Science*, Institute of Mathematical Statistics, pp. 9, 55-76.
- Hajarisman, N. (2013), *Pemodelan Area Kecil untuk Menduga Angka Kematian Bayi melalui Pendekatan Model Regresi Poisson Bayes Berhierarchy Dua Level*, Disertasi: Institut Pertanian Bogor, Bogor.

- Ha, N.S. (2013), *Hierarchical Bayes Estimation of Small Area Means Using Complex Survey Data*. Disertasi P.h.D, University of Maryland.
- Ismartini, P. (2013), Pengembangan Model Linear Hirarki dengan Pendekatan Bayesian untuk Pemodelan Data Pengeluaran Perkapita Rumah tangga, (Disertasi) Institut Teknologi Sepuluh Nopember, Surabaya.
- Jiang, J., Lahiri, P. dan Wan, S.M. (2002), *A Unified Jackknife Theory*, *Annals of Statistics*, Vol. 30, pp. 1782-1810.
- King, R., Morgan, B., Gimenez, O. dan Brooks, S. (2010), *Bayesian Analysis for Population Ecology*, USA: Chapman & Hall/CRC.
- Liu, B. (2009), *Hierarchical Bayes Estimation and Empirical Best Prediction of Small Area Proportions*, University of Maryland, Maryland: College Park.
- Lopez-Vizcaino, E., Lombardia, M. dan Morales, D. (2013), *Multinomial-Based Small Area Estimation of Labour Force Indicators*, *Statistical Modelling*, Vol. 13, pp. 153-178.
- Lopez-Vizcaino, E., Lombardia, M. dan Morales, D. (2015), *Small Area Estimation of Labour Force Indicators under Multinomial Model with Correlated Time and Area Effects*, *Journal of the Royal Statistical Society. Series A: Statistics in Society*, Vol. 178, pp. 535-565.
- Maiti, T. (1997), *Hierarchical Bayes Estimation of Mortality Rates for Disease Mapping*, *Journal of Statistical Planning and Inference*, Vol. 69, pp. 339-348.
- Miranti, P., Rumiati, A. dan Ratnasari, V. (2015), Pendugaan Tingkat Partisipasi Angkatan Kerja dan Tingkat Pengangguran Terbuka di Kabupaten Pamekasan menggunakan Small Area Estimation dengan Pendekatan *Hierarchical Bayes*, *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika 2015*, Surabaya: Universitas Negeri Surabaya, pp. 362-373.
- Molina, I., Saei, A. dan Lombardia, M. (2007), *Small Area Estimation of Labour Force Participation under Multinomial Logit Mixed Model*, *Journal of the Royal Statistical Society. Series A: Statistics in Society*, Vol. 170, pp. 975-1000.
- Noviani, A. (2016), *Small Area Estimation dengan Pendekatan Hierarchical Bayesian Neural Networks untuk Kasus Anak Putus Sekolah dari Rumah Tangga Miskin di Provinsi Jawa Timur*, Tesis. Institut Teknologi Sepuluh Nopember, Surabaya.
- Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, 1st ed., Wiley, New Jersey.
- Orme, B. (2000). *Hierarchical Bayes : Why All the Attention?* Washington: Sawtooth Software Inc.

- Peraturan Daerah Khusus Provinsi Papua Nomor 13 Tahun 2016 tentang Perubahan atas Peraturan Daerah Khusus Provinsi Papua Nomor 25 tahun 2013 tentang Pembagian Penerimaan dan Pengelolaan Keuangan Dana Otonomi Khusus.
- Pertiwi, R. dan Iriawan, N. (2012), Pemodelan Pengeluaran Per Kapita Per Kabupaten/Kota di Kalimantan Barat Menggunakan Metode Hirarki Bayesian", *IndoMS Journal on Statistics*, Vol. 20, pp. 1-14.
- Prasad, N.G.N. dan Rao, J.N.K. (1990), *The Estimation of Mean Squared Errors of Small Area Estimators*, *Journal of American Statistical Association*, Vol. 85, pp. 163-171.
- Rao, J.N.K. (2003), *Small Area Estimation*, Wiley, New York.
- Rao, J.N.K. (2007), *Jackknife and Bootstrap Methods for Small Area Estimation*, Section on Survey Research Methods.
- Rumiati, A.T. (2012), Model Bayes untuk Pendugaan Area Kecil dengan Penarikan Contoh Berpeluang Tidak Sama pada Kasus Respon Binomial dan Multinomial, Disertasi: Institut Pertanian Bogor, Bogor.
- Rumiati, A.T., Notodipuro, K.A. dan Mangku, I.W. (2012), *Empirical Bayesian Method for The Estimation of Literacy Rate at Sub-district Level. Case Study: Sumenep District of East Java Province*, IPTEK, *The Journal for Technology and Science*, pp. 23: 1-7.
- Ross, S. (2014). *Introduction to Probability Models (11th Edition)*. Oxford: Academic Press.
- Santoso, B. (2009), Pendekatan Spline Multivariable dan Mars untuk Pemodelan Lama Sekolah pada Penduduk Usia Sekolah di Provinsi Papua, Tesis: Institut Teknologi Sepuluh Nopember, Surabaya.
- Scealy, J. (2010), *Small Area Estimation Using a Multinomial Logit Mixed Model with Category Specific Random Effects*, Australian Bureau of Statistics and Australian National University, Canberra: Australian Bureau of Statistics.
- Solikhah, A. (2009), Analisis Rata-Rata Lama Sekolah di Pulau Kalimantan menggunakan Model *Spasial Conditional Autoregression (CAR)*, Tesis: Institut Teknologi Sepuluh Nopember, Surabaya.
- 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: Jurusan Ilmu Pendidikan dan Keguruan Universitas Singaraja, Bali.
- Sukarno, M. (2002), "Latar Belakang Sosial dan Pencapaian Pendidikan.", *Penduduk dan Pembangunan*, XIII (2), pp. 1-20.
- Trevisani, M. and Torelli, N. (2007), *Hierarchical Bayesian Models for Small Area Estimation with Count Data*, Università degli Studi in Trieste, Italy.

- Undang-undang No. 21 Tahun 2001 Tentang Otonomi Khusus Bagi Provinsi Papua.
- Undang-undang No. 32 Tahun 2004 Tentang Pemerintahan Daerah.
- Undang-undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional.
- UNDP (2006), *Human Development Report 2016: Overview Human Development for Everyone*, The United Development Programme, New York: USA.
- UNDP (2016), *Human Development Report 2016: Human Development for Everyone*, The United Development Programme, New York, USA.
- Wulansari, I.K. (2015), Pendugaan Statistik Area Kecil dengan Metode *Hierarchical Bayes* (HB) pada Data Respon Biner Menggunakan *Gibbs Sampling*, Tesis, Universitas Padjajaran, Bandung.
- Zhou, Q.M., dan You, Y., (2008), *Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey*, Proceedings of the Survey Methods Section, SSC Annual Meeting.

Lampiran 1. Syntax Metode *Hierarchical Bayes*

```
model;
{
  for( i in 1 : M ) {
    for( j in 1 : K ) {
      theta[i , j] <- b0[j] + b[1 , j] * x1[i] + b[2 , j] * x2[i] + b[3 , j] * x3[i] + b[4 , j] * x4[i] + b[5,j]
* x5[i] + v[i, j]
      exptheta[i , j] <- exp(theta[i , j])
      p[i , j] <- exptheta[i , j] / sum(exptheta[i , 1:K]) }
    y[i , 1:K] ~ dmulti(p[i , 1:K], n[i]) }

b[1,1] ~ dnorm(-2.0660, 76.6775)
b[1,2] ~ dnorm(0.6118, 250.2021)
b[1,3] ~ dnorm(-0.0457, 319.1054)
b[1,4] ~ dnorm(-1.4430, 401.2831)
b[1,5] ~ dnorm(-0.8657, 301.6176)
b[1,6] ~ dnorm(-2.1280, 237.2701)
b[2,1] ~ dnorm(-0.8182, 1847.2755)
b[2,2] ~ dnorm(-0.8342, 7030.1054)
b[2,3] ~ dnorm(-0.0826, 40812.1620)
b[2,4] ~ dnorm(-0.1259, 50802.6537)
b[2,5] ~ dnorm(-0.0620, 73309.9610)
b[2,6] ~ dnorm(-0.2151, 29539.4771)
b[3,1] ~ dnorm(-1.7390, 357.7965)
b[3,2] ~ dnorm(0.2058, 888.4131)
b[3,3] ~ dnorm(-1.3730, 1865.9414)
b[3,4] ~ dnorm(-1.2420, 2136.7471)
b[3,5] ~ dnorm(-1.6200, 1923.6688)
b[3,6] ~ dnorm(-2.1260, 1333.6055)
b[4,1] ~ dnorm(2.0740, 218.2913)
b[4,2] ~ dnorm(1.2040, 744.4783)
b[4,3] ~ dnorm(-0.7323, 5533.3235)
b[4,4] ~ dnorm(-0.4992, 5038.2622)
b[4,5] ~ dnorm(-0.4086, 3797.1011)
b[4,6] ~ dnorm(-0.1061, 3698.4545)
b[5,1] ~ dnorm(0.3314, 66827.0392)
b[5,2] ~ dnorm(0.3211, 387229.3315)
b[5,3] ~ dnorm(0.2709, 537359.7500)
b[5,4] ~ dnorm(0.2877, 564615.3225)
b[5,5] ~ dnorm(0.2831, 570171.3346)
b[5,6] ~ dnorm(0.2493, 496632.5828)
b0[1] ~ dnorm(-0.6943, 13.7378)
b0[2] ~ dnorm(-7.6780, 41.9583)
b0[3] ~ dnorm(-1.8430, 115.7696)
b0[4] ~ dnorm(-1.4690, 150.9957)
b0[5] ~ dnorm(-2.0500, 117.7376)
b0[6] ~ dnorm(2.2740, 74.5732)
```


Lampiran 2. Hasil Estimasi Parameter menggunakan Metode HB

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b[1,1]	-2.062	0.1092	0.001437	-2.279	-2.062	-1.851	2001	8000
b[1,2]	0.6115	0.06322	7.916E-4	0.4884	0.6125	0.7365	2001	8000
b[1,3]	-0.0464	0.05507	5.437E-4	-0.1539	-0.04653	0.0611	2001	8000
b[1,4]	-1.444	0.04872	4.625E-4	-1.537	-1.445	-1.347	2001	8000
b[1,5]	-0.8673	0.05561	7.095E-4	-0.9786	-0.8668	-0.7593	2001	8000
b[1,6]	-2.127	0.06382	7.112E-4	-2.253	-2.126	-2.004	2001	8000
b[2,1]	-0.8205	0.02093	4.361E-4	-0.8612	-0.8202	-0.7794	2001	8000
b[2,2]	-0.8342	0.01145	1.436E-4	-0.8571	-0.8341	-0.812	2001	8000
b[2,3]	-0.08261	0.004901	4.559E-5	-0.09232	-0.08264	-0.07309	2001	8000
b[2,4]	-0.1259	0.004451	5.112E-5	-0.1345	-0.126	-0.1172	2001	8000
b[2,5]	-0.06201	0.003691	4.063E-5	-0.06918	-0.06201	-0.05484	2001	8000
b[2,6]	-0.215	0.005746	6.76E-5	-0.2263	-0.2149	-0.2038	2001	8000
b[3,1]	-1.742	0.05102	6.42E-4	-1.841	-1.742	-1.642	2001	8000
b[3,2]	0.2056	0.03284	3.149E-4	0.1409	0.2053	0.271	2001	8000
b[3,3]	-1.373	0.02286	2.913E-4	-1.417	-1.373	-1.328	2001	8000
b[3,4]	-1.242	0.02123	2.259E-4	-1.283	-1.242	-1.2	2001	8000
b[3,5]	-1.62	0.02247	2.444E-4	-1.663	-1.62	-1.577	2001	8000
b[3,6]	-2.126	0.02718	3.287E-4	-2.18	-2.125	-2.072	2001	8000
b[4,1]	2.07	0.06107	0.001063	1.953	2.07	2.191	2001	8000
b[4,2]	1.204	0.03538	4.162E-4	1.135	1.204	1.275	2001	8000
b[4,3]	-0.7325	0.01311	1.572E-4	-0.7576	-0.7324	-0.7063	2001	8000
b[4,4]	-0.4991	0.01401	1.649E-4	-0.5266	-0.4991	-0.4714	2001	8000
b[4,5]	-0.4081	0.0162	1.624E-4	-0.4395	-0.4082	-0.3764	2001	8000
b[4,6]	-0.1061	0.01644	1.964E-4	-0.1383	-0.1062	-0.07414	2001	8000
b[5,1]	0.3319	0.00305	4.219E-5	0.3258	0.3319	0.3377	2001	8000
b[5,2]	0.3211	0.00156	1.669E-5	0.318	0.3211	0.3241	2001	8000
b[5,3]	0.2709	0.001299	1.406E-5	0.2684	0.2709	0.2734	2001	8000
b[5,4]	0.2877	0.001283	1.339E-5	0.2852	0.2877	0.2902	2001	8000
b[5,5]	0.2831	0.001261	1.305E-5	0.2806	0.2831	0.2856	2001	8000
b[5,6]	0.2493	0.001413	1.495E-5	0.2465	0.2493	0.2521	2001	8000
b0[1]	-0.693	0.2475	0.003731	-1.186	-0.6943	-0.2114	2001	8000
b0[2]	-7.677	0.1508	0.001718	-7.973	-7.678	-7.382	2001	8000
b0[3]	-1.843	0.09075	9.72E-4	-2.019	-1.843	-1.665	2001	8000
b0[4]	-1.47	0.08006	9.584E-4	-1.627	-1.471	-1.313	2001	8000
b0[5]	-2.049	0.09005	9.216E-4	-2.225	-2.051	-1.873	2001	8000
b0[6]	2.276	0.1128	0.00112	2.061	2.275	2.5	2001	8000
p[1,1]	0.159	0.02133	2.077E-4	0.1199	0.1587	0.2024	2001	8000
p[1,2]	0.08228	0.01604	1.578E-4	0.0537	0.08125	0.1169	2001	8000
p[1,3]	0.1625	0.02184	2.226E-4	0.1222	0.1619	0.2073	2001	8000
p[1,4]	0.22	0.02503	2.646E-4	0.1729	0.2194	0.2705	2001	8000
p[1,5]	0.2781	0.02689	3.345E-4	0.2267	0.2776	0.3322	2001	8000
p[1,6]	0.09811	0.01784	1.803E-4	0.0656	0.09739	0.1349	2001	8000
p[2,1]	1.997E-4	4.617E-4	5.434E-6	1.144E-5	1.078E-4	8.943E-4	2001	8000
p[2,2]	6.019E-4	0.001402	1.462E-5	3.375E-5	3.169E-4	0.002795	2001	8000
p[2,3]	0.5009	0.1097	0.001296	0.2895	0.5013	0.7124	2001	8000
p[2,4]	0.255	0.09269	0.001046	0.09914	0.2472	0.4579	2001	8000
p[2,5]	0.1797	0.08054	8.063E-4	0.05667	0.1688	0.3665	2001	8000
p[2,6]	0.06353	0.04482	5.402E-4	0.009865	0.05233	0.1787	2001	8000
p[3,1]	0.1778	0.05978	6.615E-4	0.07859	0.1709	0.309	2001	8000
p[3,2]	0.003109	0.003916	4.536E-5	2.387E-4	0.001907	0.01339	2001	8000
p[3,3]	0.1775	0.05874	6.179E-4	0.07886	0.172	0.3046	2001	8000
p[3,4]	0.3202	0.07472	8.393E-4	0.1831	0.3162	0.4755	2001	8000
p[3,5]	0.2669	0.07007	7.88E-4	0.1433	0.2636	0.4145	2001	8000
p[3,6]	0.05452	0.03194	3.43E-4	0.01235	0.04814	0.1366	2001	8000
p[4,1]	0.1514	0.02733	2.797E-4	0.1014	0.1499	0.2102	2001	8000
p[4,2]	0.07419	0.01965	1.956E-4	0.04023	0.0725	0.1175	2001	8000
p[4,3]	0.2056	0.03103	3.484E-4	0.1491	0.2044	0.2701	2001	8000
p[4,4]	0.2909	0.03491	4.177E-4	0.2248	0.2899	0.3602	2001	8000
p[4,5]	0.2061	0.03076	3.511E-4	0.1488	0.2047	0.2699	2001	8000
p[4,6]	0.07181	0.01932	1.922E-4	0.03915	0.06999	0.1145	2001	8000
p[5,1]	0.4499	0.05704	6.645E-4	0.3404	0.4499	0.5618	2001	8000
p[5,2]	0.09519	0.03191	3.236E-4	0.04306	0.09158	0.1672	2001	8000
p[5,3]	0.2339	0.04801	5.329E-4	0.1484	0.2317	0.3331	2001	8000
p[5,4]	0.08304	0.02949	2.996E-4	0.03561	0.07926	0.1494	2001	8000
p[5,5]	0.1114	0.03455	3.613E-4	0.05338	0.108	0.1887	2001	8000
p[5,6]	0.02657	0.0155	1.593E-4	0.005706	0.02338	0.06496	2001	8000
p[6,1]	0.3406	0.07258	6.869E-4	0.208	0.3373	0.4906	2001	8000

Lampiran 2. Hasil Estimasi Parameter menggunakan Metode HB (lanjutan)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
p[6,2]	0.1592	0.05583	6.172E-4	0.06797	0.1545	0.2807	2001	8000
p[6,3]	0.1918	0.05967	5.837E-4	0.0901	0.1869	0.3226	2001	8000
p[6,4]	0.1869	0.05972	6.0E-4	0.08576	0.1822	0.3181	2001	8000
p[6,5]	0.1187	0.04747	4.666E-4	0.04264	0.113	0.2263	2001	8000
p[6,6]	0.002701	0.003317	3.169E-5	2.025E-4	0.001646	0.01153	2001	8000
p[7,1]	0.913	0.03321	3.485E-4	0.8371	0.917	0.9656	2001	8000
p[7,2]	0.0193	0.01423	1.379E-4	0.002857	0.01567	0.05643	2001	8000
p[7,3]	0.01663	0.0125	1.318E-4	0.002463	0.01333	0.05031	2001	8000
p[7,4]	0.02063	0.01463	1.634E-4	0.003295	0.01713	0.05749	2001	8000
p[7,5]	0.03022	0.0181	2.155E-4	0.006587	0.02649	0.07554	2001	8000
p[7,6]	1.938E-4	3.095E-4	3.124E-6	1.066E-5	9.952E-5	9.731E-4	2001	8000
sigma[1]	0.9457	0.1548	0.001566	0.7018	0.9265	1.308	2001	8000
sigma[2]	0.9592	0.1575	0.001586	0.7088	0.9384	1.32	2001	8000
sigma[3]	0.9211	0.1457	0.001498	0.6915	0.9033	1.253	2001	8000
sigma[4]	0.924	0.1486	0.001577	0.6864	0.9047	1.275	2001	8000
sigma[5]	0.9232	0.1472	0.001783	0.6871	0.9069	1.267	2001	8000
sigma[6]	0.959	0.1583	0.00182	0.7071	0.9401	1.315	2001	8000
v[1,1]	-0.09569	0.6179	0.01552	-1.275	-0.09835	1.128	2001	8000
v[1,2]	0.2122	0.5383	0.01083	-0.8589	0.2056	1.275	2001	8000
v[1,3]	0.01148	0.4535	0.01008	-0.8676	0.002689	0.8923	2001	8000
v[1,4]	0.007045	0.4509	0.01007	-0.8806	0.005348	0.8882	2001	8000
v[1,5]	0.04039	0.4497	0.01021	-0.8358	0.04343	0.9296	2001	8000
v[1,6]	-0.06439	0.4769	0.01005	-1.011	-0.06868	0.8746	2001	8000
v[2,1]	-0.0255	0.9489	0.01074	-1.954	-0.01759	1.806	2001	8000
v[2,2]	-0.01458	0.9716	0.009292	-1.951	-0.02187	1.904	2001	8000
v[2,3]	0.001336	0.5821	0.006755	-1.143	0.001982	1.172	2001	8000
v[2,4]	-0.07527	0.6197	0.006261	-1.29	-0.07278	1.131	2001	8000
v[2,5]	0.09934	0.6416	0.007033	-1.183	0.1094	1.349	2001	8000
v[2,6]	-0.01684	0.7508	0.009318	-1.573	0.007579	1.393	2001	8000
v[3,1]	-0.02582	0.5846	0.007802	-1.18	-0.03012	1.086	2001	8000
v[3,2]	-0.1276	0.9257	0.01073	-1.992	-0.1077	1.665	2001	8000
v[3,3]	-0.03635	0.5448	0.006948	-1.12	-0.03416	1.026	2001	8000
v[3,4]	0.07749	0.5228	0.006701	-0.9861	0.08826	1.09	2001	8000
v[3,5]	-0.02498	0.5297	0.007016	-1.093	-0.02057	1.019	2001	8000
v[3,6]	0.1295	0.6722	0.006926	-1.246	0.1427	1.412	2001	8000
v[4,1]	-0.0527	0.5299	0.01145	-1.066	-0.05105	0.9903	2001	8000
v[4,2]	-0.2458	0.4961	0.008266	-1.225	-0.2504	0.7218	2001	8000
v[4,3]	0.03649	0.4325	0.00866	-0.8139	0.03647	0.8787	2001	8000
v[4,4]	0.05877	0.4295	0.007583	-0.7923	0.05711	0.8938	2001	8000
v[4,5]	0.05072	0.4341	0.008372	-0.8066	0.05352	0.8846	2001	8000
v[4,6]	0.1102	0.4787	0.008231	-0.8242	0.1166	1.052	2001	8000
v[5,1]	0.1886	0.5051	0.006763	-0.8056	0.187	1.175	2001	8000
v[5,2]	0.01233	0.5275	0.007317	-1.017	0.01671	1.028	2001	8000
v[5,3]	-0.07771	0.4731	0.00715	-1.018	-0.07257	0.8408	2001	8000
v[5,4]	-0.1318	0.5183	0.006617	-1.165	-0.1321	0.8653	2001	8000
v[5,5]	-0.08827	0.5034	0.006424	-1.102	-0.08039	0.8825	2001	8000
v[5,6]	0.03249	0.6538	0.0069	-1.307	0.05143	1.268	2001	8000
v[6,1]	-0.2131	0.5831	0.006923	-1.376	-0.2009	0.9135	2001	8000
v[6,2]	0.2649	0.578	0.007324	-0.893	0.2698	1.392	2001	8000
v[6,3]	0.03445	0.5525	0.00641	-1.035	0.03014	1.117	2001	8000
v[6,4]	0.2084	0.5495	0.006224	-0.8729	0.2182	1.296	2001	8000
v[6,5]	-0.2153	0.572	0.005573	-1.359	-0.2086	0.9079	2001	8000
v[6,6]	-0.08438	0.9273	0.01056	-1.956	-0.07698	1.707	2001	8000
v[7,1]	0.257	0.6563	0.008883	-1.002	0.259	1.547	2001	8000
v[7,2]	-0.1428	0.7632	0.008383	-1.675	-0.1178	1.295	2001	8000
v[7,3]	-0.0158	0.7394	0.008561	-1.526	-9.373E-4	1.42	2001	8000
v[7,4]	-0.2102	0.7418	0.008373	-1.704	-0.1913	1.199	2001	8000
v[7,5]	0.1285	0.6834	0.007821	-1.242	0.1375	1.475	2001	8000
v[7,6]	-0.02868	0.9557	0.01115	-1.884	-0.02261	1.865	2001	8000

BIOGRAFI PENULIS



Penulis dilahirkan di Kota Binjai, Provinsi Sumatera Utara pada tanggal 25 November 1981, dan putra kedua dari pasangan suami istri Bapak Pionel Sinaga dan Ibu Tudaini Sirait. Saat ini penulis telah berkeluarga dengan istri bernama dr. Evalina D. Malau dan telah dikaruniai dengan dua putri, Graleezanetha F. Sinaga dan Gwyvanetha F. Sinaga.

Riwayat pendidikan penulis diawali dari SD Methodist Indonesia (1987-1993), SMP Methodist Indonesia (1993-1996), SMUN 2 Binjai (1996-1999), Universitas Sumatera Utara (USU) Medan (1999-2002) jurusan D3 Statistika dan Universitas Cenderawasih Jayapura (2006-2008) jurusan Ekonomi Pembangunan. Setelah menamatkan pendidikan D3, penulis bekerja pada BPS Provinsi Papua. Pada tahun 2015, penulis dipercaya menjabat sebagai Kepala Seksi Sosial di BPS Kabupaten Keerom, Provinsi Papua. Pada tahun 2015, penulis memperoleh kesempatan untuk mendapatkan beasiswa dari Lembaga Pengelola Dana Pendidikan (LPDP) untuk melanjutkan studi S2 pada jurusan Statistika Fakultas Matematika, Komputasi dan Sains Data Institut Teknologi Sepuluh Nopember (ITS), Surabaya. Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini dapat menghubungi penulis melalui email fieldri.sinaga.mr@gmail.com.