PENDUGAAN PROPORSI PADA AREA KECIL

(Skripsi)

Oleh

SHELA MALINDA TAMPUBOLON

FAKULTAS MATEMATIKA DAN ILMU PEGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

ABSTRAK

PENDUGAAN PROPORSI PADA AREA KECIL

Oleh

Shela Malinda Tampubolon

Pendugaan area kecil (*Small Area Estimation*) adalah pendugaan tak langsung yang digunakan untuk menganalisis wilayah kecil dengan memanfaatkan informasi dari wilayah yang lebih besar. Salah satu metode yang dapat digunakan untuk mengestimasi dalam area kecil adalah *Hierarchical Bayes* (HB). Metode HB biasanya digunakan untuk data diskrit dengan informasi yang bertingkat. Keluarga prasejahtera merupakan bentuk data diskrit yang dapat dikelompokkan menjadi dua variabel yang bersifat kategorikal dengan menggunakan analisis kluster. Dalam penelitian ini, proporsi keluarga prasejahtera diduga menggunakan metode penduga langsung dan tak langsung. Pada pendugaan tak langsung, tidak dimungkinkan dilakukan perhitungan secara analitik dikarenakan bentuk integral multidimensional. Perhitungan secara numerik dilakukan dengan metode *Markov Chain Monte Carlo* (MCMC) dengan menggunakan software Rstudio 1.0.143. Hasil penelitian menunjukkan bahwa nilai dugaan yang dihasilkan dengan menggunakan pendugaan tak langsung memiliki nilai yang lebih kecil jika dibandingkan dengan nilai dugaan dengan menggunakan penduga langsung, artinya pendugaan tak langsung lebih baik untuk menduga nilai proporsi dari data keluarga prasejahtera.

Kata Kunci: Analisis Kluster, *Hierarchical Bayes*, Keluarga Prasejahtera, *Marcov Chain Monte Carlo*, Penduga Area Kecil (*Small Area Estimation*)

ABSTRACT

PROVIDING PROPORTION ON SMALL AREA

By

Shela Malinda Tampubolon

Small Area Estimation is an indirect prediction used to analyze small areas by using information from larger areas. The method can be used to estimate in a small area is Hierarchical Bayes (HB). The HB method is usually used for discrete data with multilevel information. Pre-prosperous families are form of discrete data that can be grouped into two categorical variables by using cluster analysis. In this study, the proportion of pre-prosperous families was expected using direct and indirect estimator methods. On indirect estimation, analytical calculation is not possible due to the multidimensional integral form. The numerical calculation is done by Markov Chain Monte Carlo (MCMC) method using Rstudio 1.0.143 software. The results showed that the estimated value generated by using indirect estimation has a smaller value than the estimated value generated by using direct estimation, it means indirect estimation is better to estimate the value of the proportion of the pre-prosperous family data.

Keywords : Cluster Analysis, Hierarchical Bayes, Pre-prosperous families, Markov Chain Monte Carlo (MCMC), Small Area Estimation

PENDUGAAN PROPORSI PADA AREA KECIL

Oleh

SHELA MALINDA TAMPUBOLON

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA SAINS

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **PENDUGAAN PROPORSI PADA AREA KECIL**

Nama Mahasiswa : Shela Malinda Tampubolon

Nomor Pokok Mahasiswa : 1317031077

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Widiarti, S.Si., M.Si
NIP. 19800502 200501 2 003

Suharsono S., M.S., M.Sc., Ph.D.
NIP.19620513 198603 1 003

2. Ketua Jurusan Matematika

Dra.Wamiliana, M.A., Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Widiarti, S.Si., M.Si.** ....

   Sekretaris : **Suharsono S., M.S., M.Sc., Ph.D.** ....

   Penguji
   Bukan Pembimbing : **Prof. Mustofa Usman, MA, Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **22 Januari 2018**

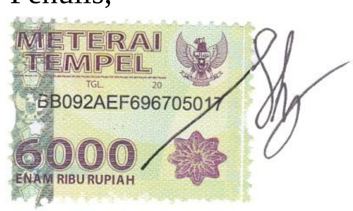
PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini :

Nama	: Shela Malinda Tampubolon
Nomor Pokok Mahasiswa	: 1317031077
Judul	: Pendugaan Proporsi pada Area Kecil
Jurusan	: Matematika

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, Januari 2018
Penulis,



Shela Malinda Tampubolon
NPM. 1317031077

RIWAYAT HIDUP

Penulis dilahirkan di Metro pada tanggal 4 September 1995. Penulis merupakan anak kedua dari pasangan Bapak Hotma Tampubolon dan Ibu Nun Helma serta adik dari Alexander Tampubolon.

Penulis memulai pendidikan dari taman kanak-kanak TK Pertiwi Teladan Metro tahun 2000. Pendidikan sekolah dasar di SD Pertiwi Teladan Metro tahun 2001. Pendidikan sekolah menengah pertama di SMP N 4 Metro pada tahun 2007. Pendidikan sekolah menengah atas di SMA Kristen 1 Metro pada tahun 2010.

Penulis melanjutkan pendidikan di perguruan tinggi dan terdaftar sebagai mahasiswa Jurusan Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN pada tahun 2013. Pada periode tahun 2014/2015 penulis terdaftar sebagai anggota bidang Keilmuan Himpunan Mahasiswa Matematika FMIPA Unila dan anggota bidang Redaksi UKMF Natural FMIPA Unila. Penulis juga mendapatkan amanah menjadi Redaktur Media Cetak bidang Redaksi pada periode 2015/2016 dan menjadi Pemimpin Redaksi UKMF Natural FMIPA Unila pada periode 2016/2017.

Sebagai bentuk aplikasi bidang ilmu kepada masyarakat, penulis telah menyelesaikan Kerja Praktik (KP) di kantor BPS Kota Metro selama kurang lebih dua minggu. Penulis juga telah melakukan Kuliah Kerja Nyata (KKN) pada tahun 2016 selama 60 hari di Desa Gedung Bandar Rahayu, Kecamatan Gedung Meneng, Kabupaten Tulang Bawang, Provinsi Lampung.

MOTTO

“Orang yang menuntut ilmu bearti menuntut rahmat ; orang yang menuntut ilmu bearti menjalankan rukun Islam dan Pahala yang diberikan kepada sama dengan para Nabi”.

(HR. Dailani dari Anas r.a)

"Orang-orang yang sukses telah belajar membuat diri mereka melakukan hal yang harus dikerjakan ketika hal itu memang harus dikerjakan, entah mereka menyukainya atau tidak."

(Aldus Huxley)

PERSEMBAHAN

Sembah sujud serta syukur saya ucapkan kepada Allah SWT. Taburan cinta serta kasih sayang-Nya telah memberikanku kekuatan dan membekaliku dengan ilmu yang bermanfaat. Dengan kemurahan-Nya, akhirnya skripsi yang sederhana ini dapat terselesaikan. Sholawat dan salam selalu terlimpahkan ke Nabi besar kita Rasulullah Muhammad SAW.

Aku persembahkan skripsi sederhana ini untuk keluarga yang kusayangi. Untuk Bapak dan Mama yang tidak pernah putus memberikan kasih sayang dan dukungan meskipun berat sekalipun, kalian adalah *Superhero* terhebat yang pernah ada.

Untuk Abang Alex dan Mbak Wiwid, terima kasih untuk perhatian yang tak kurang-kurangnya. Walaupun ngeselin, tapi adek ucapin banyak terima kasih karena udah ngerti posisi adek waktu ngerjain skripsi ini.

Buat Bang Angga dan Ence, terima kasih banyak udah mau nerima aku di rumah kalian. Kalian udah kayak abang dan adek kandung sendiri yang selalu kasih semangat dan asupan kebahagiaan lewat apapun itu.

For everyone, I can’t say your name one by one, I do thank to you. For everything you gave to me, I can’t give it back as much as you gave, but let Allah SWT gives all best things for you. Amin.

SANWACANA

Puji syukur kehadirat Allah SWT berkat rahmat dan hidayahnya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Shalawat serta salam senantiasa tercurah kepada junjungan kita Nabi Muhammad SAW, suri tauladan terbaik sepanjang masa.

Pada proses penyusunan skripsi, penulis memperoleh banyak bantuan, dukungan, bimbingan serta kritik dan saran yang membangun sehingga skripsi ini mampu terselesaikan. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Ibu Widiarti, S.Si., M.Si., selaku dosen pembimbing utama dan ibu yang mengasihi anak bimbingannya, terima kasih banyak ibu atas ilmu dan kasih sayang ibu, begitu banyak rasa terima kasih penulis kepada ibu yang telah memotivasi dan mengajarkan banyak hal diluar perkuliahan di kampus.
2. Bapak Suharsono S., M.S., M.Sc., Ph.D., selaku dosen pembimbing pembantu yang telah banyak membantu dalam proses penyusunan skripsi ini.
3. Bapak Drs. Mustofa Usman, MA, Ph.D., selaku dosen penguji yang telah memberikan kritik dan saran yang membangun dalam proses penyelesaian skripsi.
4. Ibu Dr. Asmiati, S.Si., M.Si., selaku Pembimbing Akademik yang selalu memberikan motivasi kepada penulis.
5. Ibu Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika FMIPA Universitas Lampung yang telah banyak memberikan dukungan kepada penulis.

6. Bapak Prof. Warsito, S.Si., D.E.A, Ph.D., selaku Dekan FMIPA Universitas Lampung.
7. Dosen, staf dan karyawan Jurusan Matematika FMIPA UNILA yang telah memberikan ilmu pengetahuan dan segala bentuk bantuan kepada penulis.
8. Bapak Hotma, Mama Nun, Abang Alex, Mbak Wiwid, Abang Angga, Ence, dan Keluarga Besar yang telah memberikan banyak dukungan, kasih sayang dan motivasi kepada penulis setiap waktunya.
9. Teman-teman satu pembimbing yaitu Della, Rifa, Nina, Galuh, Dita, Lia, Chaterine, terima kasih untuk dukungannya kawan. Semoga kita tetap menjalin silaturahim tanpa terputus oleh jarak dan waktu.
10. Sosok yang selalu ada untuk penulis, Cinkia Eagseli Ewys, Tasya Marina, Retno Safitri, Nurul Nikmah, Yeftanus Antonio, Nandra Adi P, Ranu Solehiman, Mbak Lia, terima kasih banyak untuk semangat yang kalian selalu berikan, semoga kita terus menjalin komunikasi dan menjaga baik hubungan ini.
11. Keluarga besar Matematika 2013, HIMATIKA dan NATURAL yang telah memberikan banyak ilmu yang sangat berharga dan mengajarkan arti keluarga tak sedarah.
12. Seluruh pihak yang telah membantu penulis yang tidak dapat disebutkan satu persatu atas peran dan dukungannya dalam menyusun skripsi ini.

Bandar lampung,   Januari 2018
Penulis,

Shela Malinda Tampubolon

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Survei merupakan salah satu alat bantu bagi masyarakat untuk memperoleh suatu informasi dari suatu wilayah dan menjadi salah satu bagian yang penting dari suatu proses pengambilan data. Umumnya survei dilakukan untuk mengumpulkan informasi untuk area kecil seperti kecamatan atau kelurahan. Namun, terkadang dalam mengumpulkan informasi ditemukan beberapa kendala seperti ketidaksediaan responden dalam mengisi kuisioner, responden yang tidak ada ditempat, dan lain-lain. Data hasil survei tersebut nantinya akan digunakan untuk mengestimasi area kecil.

Pendugaan area kecil (*Small Area Estimation*) adalah suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil. Teknik pendugaan ini memanfaatkan data dari domain besar untuk menduga peubah yang menjadi perhatian pada domain yang lebih kecil. Area kecil didefinisikan sebagai subpopulasi yang ukuran contohnya kecil sehingga pendugaan langsung tidak dapat menghasilkan dugaan yang teliti (Rao, 2003).

Pendugaan langsung (*direct estimation*) dilakukan berdasarkan data sampel yang tersedia, pendugaan yang dihasilkan merupakan penduga tak bias, tetapi memiliki varian yang besar karena diperoleh dari ukuran sampel yang kecil. Pendugaan tak langsung (*indirect estimation*) merupakan suatu pendugaan dengan cara memanfaatkan atau meminjam informasi tambahan yang berhubungan dengan parameter yang akan diamati.

Terdapat beberapa model estimasi dalam area kecil yang dapat digunakan, seperti *Best Linear Unbiased Predictor* (BLUP), *Empirical Best Linear Unbiased Predictor* (EBLUP), *Empirical Bayes* (EB), dan *Hierarchical Bayes* (HB). Untuk mendapatkan suatu estimasi yang baik, diperlukan suatu model yang mampu menggabungkan informasi baik dari dalam wilayah maupun luar wilayah dengan menghitung prior dan posterior.

Hierarchical Bayes (HB) atau Bayes Hirarki merupakan salah satu metode Bayes yang dapat digunakan untuk menduga variabel independen dari data diskrit. Metode HB dapat digunakan dalam mengatasi model dengan kelas distribusi selain distribusi normal dan memperkirakan parameter dari distribusi posterior dengan menggunakan metode Bayes. Model Hirarki akan terbentuk melalui sub model yang bergabung dan akan berintegrasi dengan data yang diamati dengan menggunakan teorema Bayes. Selain itu, metode HB juga dapat digunakan ketika informasi yang tersedia pada beberapa tingkat berbeda dari unit pengamatan. Metode HB dinilai lebih menguntungkan karena mempunyai kuadrat tengah galat yang minimum.

Data biner merupakan data dengan variabel dikotomi yang hanya memiliki dua kategori dimana disimbolkan dengan nilai 0 dan 1. Angka 0 dan 1 merupakan simbol dari suatu kejadian tertentu, seperti gagal dan berhasil. Salah satu contohnya yaitu data keluarga sejahtera. Untuk kasus dengan data biner, variabel yang menjadi perhatian berupa proporsi.

Hubungan antara variabel dari data biner dapat dijelaskan dengan metode logistik. Metode logistik merupakan sebuah pendekatan untuk membuat model prediksi variabel terikat yang berskala dikotomi dengan variabel bebas yang berupa data berskala interval dan atau kategorik. Hal ini dilakukan karena data tidak diwajibkan untuk memenuhi asumsi sebaran normal dengan ragam yang homogen.

Pada penelitian ini penulis tertarik untuk mengkaji penduga proporsi dengan metode HB dan membandingkan penduga proporsi menggunakan metode HB dengan penduga langsung.

1.2 Tujuan Penelitian

Penelitian ini memiliki tujuan :

1. Mengaplikasikan metode HB pada pendugaan proporsi keluarga prasejahtera
2. Membandingkan penduga proporsi dengan metode HB dan penduga langsung

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah mengetahui perbandingan pendugaan proporsi dengan metode HB.

II. TINJAUAN PUSTAKA

2.1 *Small Area Estimation/SAE* (Pendugaan Area Kecil)

Pendugaan area kecil merupakan metode estimasi tidak langsung yang mengombinasikan antara data survei dengan data pendukung lain, misalnya dari data sensus sebelumnya yang memuat variabel dengan karakteristik yang sama dengan data survei sehingga dapat digunakan untuk menduga area yang lebih kecil dan memberikan tingkat akurasi yang lebih baik. Model area kecil dikelompokkan menjadi dua jenis yaitu:

1. Pendugaan Area Kecil Berbasis Area

Pada model pendugaan area kecil berbasis area, data pendukung yang tersedia hanya sampai level area. Misalkan $x_i = (x_{1i}, x_{2i}, \dots x_{pi})^T$ dengan parameter yang akan diduga adalah θ_i yang diasumsikan mempunyai hubungan dengan y_i. Data pendukung tersebut digunakan untuk membangun model :

$$\hat{\theta}_i = x_i^T \beta + z_i v_i \quad , \quad i = 1, \dots, m \qquad (2.1)$$

Di mana *m* adalah jumlah daerah dengan β merupakan vektor koefisien regresi untuk data pendukung x_i dan v_i berdistribusi independen $N(0, \sigma_v^2)$ sebagai pengaruh acak area spesifik yang diasumsikan berdistribusi normal.

Estimator θ_i dapat diketahui dengan mengasumsikan bahwa model estimator langsung $\widehat{\theta}_i$ telah tersedia, yaitu :

$$\widehat{\theta}_i = \theta_i + e_i \quad , \quad i = 1, \dots, m \tag{2.2}$$

Dengan $e_i \sim N(0, \sigma_\varepsilon^2)$ dan σ_ε^2 diketahui. Jika kedua model digabungkan maka akan menghasilkan model gabungan (*mixed model*) :

$$\widehat{\theta}_i = x_i^T \beta + z_i v_i + e_i \quad , \quad i = 1, \dots, m \tag{2.3}$$

(Rao, 2003).

2. Pendugaan Area Kecil Berbasis Unit

Pada model pendugaan area kecil berbasis unit diasumsikan bahwa data variabel penyerta unit $x_i = (x_{ij1}, x_{ij2}, \dots x_{ijp})^T$ tersedia untuk setiap elemen ke-j pada area ke-i, sehingga dapat dibangun suatu model regresi tersarang :

$$y_{ij} = x_{ij}^T \beta + z_i v_i + e_{ij} \quad , \quad i = 1, \dots, m \; dan \; j = 1, \dots, N \tag{2.4}$$

Dengan $v_i \sim N\left(0, \sigma_v^2\right)$ dan $e_i \sim N(0, \sigma_\varepsilon^2)$ (Rao, 2003).

2.2 Model *Hierarchical Bayes* (HB)

Dalam HB, estimasi parameter area kecil ditaksir melalui sebaran posterior $f(\mu|y)$. Langkah pertama yang dilakukan yaitu dengan menentukan sebaran prior subjektif $f(\theta)$ pada parameter model θ, kemudian untuk gugus data y tertentu akan diperoleh sebaran posterior $f(\mu|y)$ dari parameter area kecil μ yang diamati. Model dua tahap, katakan $f(y|\mu,\theta_1)$ dan $f(y|\mu,\theta_2)$ dikombinasikan dengan prior subjektif pada $\theta = (\theta_1^T, \theta_2^T)$ dengan menggunakan teorema Bayes, sehingga diperoleh posterior $f(\mu|y)$. Inferensi yang berdasarkan pada $f(\mu|y)$, secara khusus pada suatu parameter yang diamati, sebut saja dengan $\emptyset = h(u)$, diduga dengan menggunakan rata-rata posteriornya sebagaimana ditunjukkan dalam persamaan :

$$E_{\theta|y}(\theta) = \mu(y) = \mu \quad (2.5)$$

Serta ragam posterior yang dinyatakan dalam persamaan :

$$V_{\theta|y}(\theta) = E_{\theta|y}[\theta - E_{\theta|y}(\theta)]^2 \quad (2.6)$$

(Shobri dkk, 2015)

2.3 *Markov Chain Monte Carlo* (MCMC)

Pada penggunaan metode Bayesian, ada kalanya dihadapkan pada kondisi dimana penentuan distribusi posterior sulit dilakukan karena melibatkan persamaan integral yang sangat kompleks. Misalnya pada model yang kompleks seperti model hirarki dengan banyak parameter, maka untuk mendapatkan distribusi posterior parameter diperlukan proses integral dengan dimensi yang besar dan waktu yang cukup lama.

Salah satu solusi untuk mengatasi masalah ini adalah dengan pendekatan numerik, yaitu *Markov Chain Monte Carlo* (Carlin dan Chib, 1995).

Markov Chain Monte Carlo (MCMC) adalah sebuah metode untuk membangkitkan variabel-variabel yang didasarkan pada rantai markov. Metode MCMC akan memperoleh sebuah barisan sampel acak yang berkorelasi, yakni nilai ke-j dari barisan $\{\theta_j\}$ disampling dari sebuah distribusi peluang yang bergantung pada nilai sebelumnya $\{\theta_{j-1}\}$. Distribusi eksak dari $\{\theta_j\}$ umumnya tidak diketahui namun distibusi pada setiap iterasi dalam barisan nilai sampel tersebut akan konvergen pada distibusi yang sesungguhnya untuk nilai j yang cukup besar.

Oleh karena itu, jika ukuran sampel yang diperbaharui cukup besar maka kelompok terakhir dari nilai yang disampling dalam barisan tersebut, misal $\{\theta_{p+1}, \theta_{p+2}, \ldots\}$ akan mendekati sebuah sampel yang berasal dari distribusi yang diinginkan. Notasi P biasanya disebut sebagai *burn in period*.

Terdapat dua algoritma utama dalam MCMC, yaitu algoritma Metropolis-Hastings dan algoritma *Gibbs Sampling*.

2.3.1 Algoritma Metropolis-Hasting (MH)

Algoritma Metropolis-Hasting (MH) digunakan untuk membantu membangkitkan sampel-sampel acak dari distribusi posterior yang diinginkan. Dalam algoritma MH dibutuhkan sebuah *proposal distribution* $p(\theta|\theta_{j-1})$ untuk membangkitkan kandidat sampel acak.

Langkah-langkah dasar dari algoritma ini adalah :

Langkah 1 : Ambil nilai awal, yaitu θ_0 . Untuk iterasi j=1. Bangkitkan $\theta \sim p\left(\theta|\theta_{j-1}\right)$

Langkah 2 : Bangkitkan sampel acak u dari distribusi uniform U[0,1]

Langkah 3 : Jika $u < min\left(1, \frac{p(\theta^*|X,y)p(\theta_{j-1}|\theta^*)}{p(\theta_{j-1}|X,y)p(\theta^*|\theta_{j-1})}\right)$ ambil $\theta_j = \theta^*$.

Namun, jika $u < min\left(1, \frac{p(\theta^*|X,y)p(\theta_{j-1}|\theta^*)}{p(\theta_{j-1}|X,y)p(\theta^*|\theta_{j-1})}\right)$, ambil $\theta_j = \theta_{j-1}$

Langkah 4 : Ulangi langkah 1-3 sampai dengan jumlah yang diinginkan.

Statistik yang digunakan untuk mengukur derajat ketergantungan antar pengambilan secara berurutan dalam sebuah rantai markov adalah autokorelasi. Autokorelasi mengukur korelasi antara dua buah himpunan dari nilai-nilai yang tersimulasi $\{\theta_j\}$ dan $\{\theta_{j+L}\}$ dengan L adalah lag atau bilangan yang memisahkan dua himpunan tersebut. Untuk sebuah hyperparameter θ_j tertentu, nilai autokorelasi pada lag ke-*l* dapa dihitung dengan formula :

$$r_{iL} = \left(\frac{M}{M-L}\right)\left(\frac{\sum_{i=1}^{M-L}(\theta_j - \overline{\theta})(\theta_{i+L} - \overline{\theta})}{\sum_{i=1}^{M}\left(\theta_i - \overline{\theta}\right)^2}\right) \tag{2.7}$$

dengan M adalah ukuran sampel acaknya.

2.3.2 Algoritma *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 acak dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella dan George,1992).

Gibbs Sampling bisa diterapkan apabila distribusi probabilitas bersama (*joint probability distribution*) tidak diketahui secara eksplisit, tetapi distribusi bersyarat (*conditional distribution*) dari tiap-tiap variabel diketahui. Algoritma *Gibbs Sampling* bisa dituliskan sebagai berikut :

Langkah 1 : Tentukan nilai awal $x^0 = (x_1^{(0)}, \dots, x_p^{(0)})$

Langkah 2 : Ulangi Langkah 2 untuk j=1, 2, …, m

Bangkitkan $X_1^{(j+1)}$ dari $f_1(x_1|x_2^{(j)}, x_3^{(j)}, \dots, x_p^{(j)})$

Bangkitkan $X_2^{(j+1)}$ dari $f_2(x_1|x_2^{(j)}, x_3^{(j)}, \dots, x_p^{(j)})$

Bangkitkan $X_p^{(j+1)}$ dari $f_p(x_1|x_2^{(j)}, x_3^{(j)}, \dots, x_p^{(j)})$

Langkah 3 : Kembalikan nilai $\{x^1, x^2, \dots, x^m\}$

Fungsi kepekatan $f_1, f_2, \dots, f_p$ disebut distribusi bersyarat penuh, dan fungsi kepekatan yang digunakan untuk simulasi. Walaupun dalam dimensi tinggi, semua simulasi adalah univariat. Dalam *Gibbs Sampling* tidak ada mekanisme penerimaan dan penolakan semua sampel hasil simulasi diterima.

2.4 Model HB untuk Data Biner

Rao (2003) menggunakan pendekatan HB dengan model logit-normal dan variabel berbasis area dengan model sebagai berikut :

$$\begin{aligned} &\text{I.} \quad y_i|p_i \sim ind\ Binomial\ (n_i, p_i) \\ &\text{II.} \quad \theta = logit(p_i) = X_i'\beta + v_i,\ v_i \sim N_m(0, \sigma_v^2) \\ &\text{III.} \quad \beta\ dan\ \sigma_v^2\ saling\ bebas, f(\beta) \propto 1 \\ &\qquad \sigma_v^{-2} \sim gamma(a,b); a \geq 0, b > 0 \end{aligned} \tag{2.8}$$

Pendugaan bayes dengan HB tidak dapat diselesaikan dengan analitik. Untuk mempermudah hal itu dibutuhkan pendekatan numerik, yaitu model *Marcov Chain Monte Carlo* (MCMC). Prosedur MCMC yang terkenal yaitu Gibbs bersyarat (*Gibbs Conditionals*).

Menurut Sunandi, dkk (2011), bentuk Gibbs bersyarat untuk model logit normal dengan variabel bebas berbasis area adalah :

$$\begin{aligned} &\text{i.} \quad [\beta|p, \sigma_v^2, y] \sim N_p\left[\beta^*, \sigma_v^2\left(\textstyle\sum_{i=1}^m x_i x_i'\right)^{-1}\right] \\ &\text{ii.} \quad [\sigma_v^2|\beta, p, y] \sim gamma\left[\frac{m}{2} + a, \frac{1}{2}\textstyle\sum_{i=1}^m \left(\theta_i - x_i'\beta\right)^2 + b\right] \\ &\text{iii.} \quad f(p_i|\beta, \sigma_v^2, y) \propto h(p_i|\beta, \sigma_v^2)k(p_i) \end{aligned} \tag{2.9}$$

Pendugaan parameter β dan σ_v^2 dibangkitkan secara langsung dari (i) dan (ii). Parameter β^* pada bagian (i) dinyatakan oleh :

$$\beta^* = \left(\sum_{i=1}^{m} x_i' x_i\right)^{-1} \left(\sum_{i=1}^{m} x_i' \theta_i\right) \tag{2.10}$$

Sementara itu, pada bagian (iii) dinyatakan sebagai :

1. $h(p_i|\beta, \sigma_v^2) = \frac{\partial\theta_i}{\partial p_i} \exp\left\{-\frac{1}{2\sigma_v^2}[\theta_i - x_i'\beta]^2\right\}$
2. $k(p_i) = p_i^{y_i}(1-p_i)^{n-y_i}$

(2.11)

Nilai proporsi HB akan diduga melalui simulasi *Gibbs Sampling* Metropolis-Hasting (M-H). Sampel Gibbs MCMC dapat dibangkitkan langsung dari (2). Adapun logaritma M-H adalah sebagai berikut :

1. Dibangkitkan $\theta \sim ind\ N(X'\beta, \sigma_v^2)$ lalu dicari nilai $p_i^* = g^{-1}(\theta_i)$
2. Dihitung dengan $r\left(p_i^{(k)}, p_i^*\right) = min\left\{{k(p_i^*)}\big/{k(p_i^{(k)})}, 1\right\}$; $k = 0,1, \dots, D$
3. Dibangkitkan *u* dari distribusi seragam (0,1)
4. Dipilih $p_i^{(k+1)} = p_i^*$ jika $u \leq r\left(p_i^{(k)}, p_i^*\right)$
5. Diulangi langkah 3 sampai diperoleh D sampel

Setelah dilakukan simulasi M-H, maka diperoleh barisan penduga proporsi sebagai berikut :

$$\left\{p_i^{(k)}, \dots, p_m^{(k)}; k = 1, \dots, D\right\} \quad (2.12)$$

Kemudian besaran posterior yang sedang diamati dapat dihitung. Penduga proporsi HB (p_i^{HB}) adalah :

$$p_i^{HB} \approx \frac{1}{D}\sum_{k=d+1}^{d+D} p_i^{(k)} = p_i^{(.)} \quad (2.13)$$

Sedangkan ragam penduga proporsi HB ($V(p_i^{HB}|\hat{p})$) adalah :

$$V(p_i^{HB}|\hat{p}) = \frac{1}{D}\sum_{k=d+1}^{d+D}\left(p_i^{(k)} - p_i^{(.)}\right)^2 \qquad (2.14)$$

2.5 Model Logistik/Logit

Model logit merupakan model regresi non-linear dengan persamaan yang variabelnya bersifat kategorikal. Kategori tersebut menghasilkan nilai biner yang disimbolkan dengan angka 0 dan 1. Angka 0 dan 1 menyimbolkan suatu kategori tertentu yang dari perhitungan probabilitas yang dilakukan.

Tabel 2.1 Probabilitas Model Logit

Yi	**Probabilitas**
0	1-Pi
1	Pi
Total	1

Penggunaan model logit biasanya digunakan pada data berbentuk klasifikasi. Salah satu contoh data klasifikasi adalah keluarga sejahtera, yang mana nilai 0 menyimbolkan keluarga yang sejahtera, sedangkan angka 1 menyimbolkan keluarga prasejahtera. Dalam menentukan apakah sebuah keluarga tergolong dalam keluarga sejahtera, tentunya dipengaruhi oleh beberapa variabel independen. Variabel independen tersebut dapat berupa data ordinal, nominal, rasio, dan interval.

2.6 Penduga Langsung untuk Proporsi

Variabel dari suatu pengamatan diasumsikan mempunya sebaran binomial $y_i \overset{iid}{\sim} binomial\ (n_i, p_i)$. Fungsi peluang dari sebaran binomial adalah :

$$f(y_i|p_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i} \qquad (2.15)$$

Dengan

$$y_i = 0, 1, \dots, n_i\ , 0 < p_i < 1\ , i = 1, 2, \dots, m$$

Selanjutnya dengan memaksimumkan fungsi peluang tersebut, maka diperoleh *Maximum Likelihood Estimation* (penduga kemungkinan maksimum) bagi p_i yaitu :

$$\hat{p}_i = \frac{y_i}{n_i} \qquad (2.16)$$

Penduga ini merupakan penduga kemungkinan maksimum yang bersifat tak bias karena nilai harapan dari penduga sama dengan parameternya.

$$E(\hat{p}_i) = E\left(\frac{y_i}{n_i}\right) = \frac{1}{n_i}E(y_i) = \frac{1}{n_i}n_i p_i = p_i \qquad (2.17)$$

Sehingga dugaan *Mean Square Error/MSE* (Kuadrat Tengah Galat/KTG) sama dengan ragamnya, yaitu :

$$ktg(\hat{p}_i) = \hat{V}ar(\hat{p}_i) = \frac{\hat{p}_i(1-\hat{p}_i)}{n_i} \qquad (2.18)$$

(Sahara, 2012).

2.7 Keluarga Sejahtera

Keluarga sejahtera adalah keluarga yang dibentuk berdasarkan atas perkawinan yang sah, mampu memenuhi kebutuhan hidup spiritual dan materiil yang layak, bertaqwa kepada Tuhan Yang Maha Esa, memiliki hubungan yang serasi, selaras dan seimbang antar anggota dan antar keluarga dengan masyarakat dan lingkungan (Undang-Undang Republik Indonesia Nomor 52 tahun 2009).

2.7.1 Tahapan Keluarga Pra Sejahtera (KPS)

KPS adalah keluarga yang tidak memenuhi salah satu dari enam indikator KS I atau indikator “kebutuhan dasar keluarga”.

Ada enam indikator dalam tahapan KS I atau Indikator “kebutuhan dasar keluarga”, yaitu :

1. Keluarga membeli satu stel pakaian baru untuk seluruh anggota keluarga minimal setahun sekali,
2. Seluruh anggota keluarga makan minimal 2 kali sehari,
3. Seluruh anggota keluarga bila sakit berobat ke fasilitas kesehatan,
4. Seluruh anggota keluarga memiliki pakaian yang berbeda untuk di rumah, bekerja/sekolah dan bepergian,
5. Seluruh anggota keluarga makan daging/ikan/telur minimal seminggu sekali, dan
6. Seluruh anggota keluarga menjalankan ibadah agama sesuai ketentuan agama yang dianut (BKKBN, ___).

III. METODE PENELITIAN

3. 1. Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2017/2018 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3. 2. Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder, yaitu data Keluarga Prasejahtera tahun 2015 di Kota Bandar Lampung yang diperoleh dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Provinsi Lampung.

3. 3. Metode Penelitian

Pada penelitian ini dilakukan perhitungan pendugaan proporsi keluarga prasejahtera dengan menggunakan dua metode, yaitu pendugaan langsung dan pendugaan tak langsung. Pendugaan Langsung dilakukan dengan menghitung proporsi keluarga prasejahtera, sedangkan pendugaan tak langsung dilakukan dengan metode HB.

A. Menentukan Penduga Langsung :
   1. Menentukan penduga proporsi

      $$\hat{p}_i^{DE} = \frac{y_i}{n_i}$$

dengan y_i menyatakan banyaknya pengamatan suatu kasus pada subpopulasi ke-i , n_i menyatakan banyaknya individu pada subpopulasi ke-i. Subpopulasi ini dapat berupa kecamatan.

2. Menentukan ragam proporsi

$\hat{V}(p_i) = \frac{p_i q_i}{n_i} \quad ; \quad q_i = 1 - p_i$

Dengan n_i merupakan jumlah sampel di kecamatan-i.

B. Pendugaan tak langsung dengan metode HB dilakukan dengan tahapan sebagai berikut :

1. Menetapkan model dalam HB sebagai berikut :
   - i. $y_i | p_i \overset{iid}{\sim} Binomial\ (n_i, p_i)$
   - ii. $\theta = logit(p_i) = X_i'\beta + v_i,\ v_i \sim N_m(0, \sigma_v^2)$
   - iii. $\beta\ dan\ \sigma_v^2\ saling\ bebas, f(\beta) \propto 1$

     $\sigma_v^{-2} \sim gamma(a, b); a \geq 0, b > 0$
2. Menentukan fungsi kepekatan peluang bersama model pada poin 1
3. Menentukan fungsi kepekatan peluang akhir (posterior) model pada poin 1
4. Menentukan rata-rata posterior (penduga HB) secara analitik, jika tidak dapat diselesaikan maka menggunakan metode *Gibbs Sampling*.

   Sebelum masuk ke dalam metode Gibbs Sampling, kita akan menentukan variabel dikotomi untuk proporsi keluarga pra sejahtera berdasarkan variabel penyerta dengan menggunakan Analisis Kluster. Analisis kluster adalah analisis yang digunakan untuk mengelompokkan area berdasarkan kesamaan tertentu. Dengan menggunakan aglomerasi, semakin mirip antar area, maka akan membentuk kelompok kluster tertentu. Aglomerasi adalah proses yang menganggap seluruh area sebagai satu kelompok, kemudian melihat kemiripan antar area dengan menggunakan jarak Euclidean, kemudian kluster dengan kesamaan terdekat akan digabungkan menjadi satu kluster.

Berikut adalah langkah dalam menggunakan metode *Gibbs Sampling* :

a. Mengambil nilai sembarang awal dari $p^{(0)}, \sigma_v^{2(0)}, y^{(0)}$

b. Membangkitkan $\beta^{(1)}$ dengan informasi $p^{(0)}, \sigma_v^{2(0)}, y^{(1)}$ dari $(\beta | p, \sigma_v^2, y) \sim N_p(\beta^*, \sigma_v^2 \left(\sum_{i=1}^{m} \sigma_v^2 \left(x_i x_i'\right)^{-1}\right))$, informasi $\boldsymbol{x}$ merupakan gugus variabel prediktor

c. Melakukan iterasi ke-k, akan dibangkitkan contoh acak $\beta^{(k)}$ dengan informasi $p^{(k-1)}, \sigma_v^{2(k-1)}, y^{(k)}$

d. Membangkitkan contoh acak $\sigma_v^{2(k)}$ dengan informasi $\beta^{(k)}, p^{(k-1)}, y^{(k)}$

e. Menghitung nilai $p^{(k)}$ dengan informasi $\sigma_v^{2(k)}, \beta^{(k)}, y^{(k)}$

f. Mengulangi proses sampai sejumlah D contoh acak/iterasi yang telah ditetapkan sampai rantai konvergen. Semakin banyak jumlah iterasi yang dilakukan nilai estimasi yang diperoleh akan konvergen ke suatu nilai yang mendekati nilai pada keadaan yang sebenarnya. Kekonvergenan suatu nilai juga dapat dilihat dari hasil *output* yang dikeluarkan.

g. Melakukan "*burn in*" dengan cara membuang d iterasi pertama untuk menghilangkan pengaruh nilai awal sehingga diperoleh $\left\{p_i^{(k)}, \dots, p_m^{(k)}, \beta^{(k)}, \sigma_v^{2(k)}; k = d+1, \dots, K = d + D\right\}$

h. Saat rantai konvergen, maka diperoleh nilai $\hat{p}_i^{HB}\ dan\ V(\hat{p}_i^{HB})$

5. Membandingkan $\hat{p}_i^{HB}$ dengan hasil estimasi langsung $\hat{p}_i^{DE}$

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka dapat kita ambil kesimpulan :

1. Pendugaan tak langsung dengan pendekatan HB model logit-normal dapat digunakan untuk data keluarga prasejahtera Kota Bandar Lampung tahun 2015.
2. Berdasarkan hasil perbandingan antara penduga langsung dan tak langsung, penduga tak langsung lebih baik jika dibandingkan dengan penduga langsung, karena penduga tak langsung merupakan penduga yang baik, karena penduga tak langsung memiliki nilai varian yang kecil. Untuk penduga tak langsung didapatkan nilai proporsi terbesar yaitu 0.38852534, artinya proporsi keluarga prasejahtera terbesar yaitu 38,85% berada di kecamatan Telukbetung Timur. Sedangkan nilai proporsi terkecil yaitu 0.03766538, artinya proporsi keluarga prasejahtera terkecil yaitu 3,76% berada di kecamatan Enggal. Rata-rata proporsi keluarga prasejahtera di Kota Bandar Lampung yaitu sebesar 0.11985 atau sebesar 11,985%.

DAFTAR PUSTAKA

BKKBN. ____. http://aplikasi.bkkbn.go.id/mdk/BatasanMDK.aspx, 4 Mei 2017

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

Casella, G., dan George, E. I. 1992. *Explaining the Gibbs Sampler. The American Statistician*, 46(3), 167-174.

Gelman, A., dkk. 2014. *Bayesian Data Analysis, Third Edition*. Boca Raton, FL, USA: Chapman & Hall/CRC

Rao, J.N.K. 2003. *Small Area Estimation*. John Willey and Sons, New York.

Sahara, Linda. 2012. *Penerapan Metode Bayes Empirik pada Pendugaan Area Kecil (Studi tentang Proporsi Status Kepemilikan Kartu Jaminan Kesehatan Masyarakat (JAMKESMAS) di Kota Takengon Kabupaten Aceh Tengah)* (Skripsi). Medan: Universitas Sumatera Utara.

Shobri, Ari, Padmadisastra, Septiadi, dan Winarni, Sri. 2015. *Pendekatan Hierarchical Bayes Small Area Estimation (HB SAE) dalam Mengestimasi Angka Melek Huruf Kecamatan di Kabupaten Indramayu* (Tesis). Bandung: Universitas Padjajaran.

Sunandi, K.A. Notodiputro, A. Djuraidah. 2011. *Model Spasial Bayes dalam Pendugaan Area Kecil dengan Peubah Respon Biner (Kasus : Pendugaan Proporsi Keluarga Miskin di Kabupaten Jember Jawa Timur)* (Tesis). Bogor: Program Pascasarjana, Institut Pertanian Bogor.