SKRIPSI

PENDUGAAN AREA KECIL DENGAN PENDEKATAN BAYESIAN HIERARKI

Disusun dan diajukan oleh

SITI HAJRIAH

H12116009

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2020

PENDUGAAN AREA KECIL DENGAN PENDEKATAN BAYESIAN HIERARKI

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

SITI HAJRIAH

H 121 16 009

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2020

LEMBAR PENGESAHAN (TUGAS AKHIR)

PENDUGAAN AREA KECIL DENGAN PENDEKATAN BAYESIAN HIERARKI

Disusun dan diajukan oleh

SITI HAJRIAH
H 121 16 009

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 8 Desember 2020
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Dr.Dr. Georgina Maria Tinungki, M.Si	Anisa, S.Si, M.Si
NIP. 19620926 198702 2001	NIP. 19730227 199802 2001

Ketua Progam Studi,

Dr. Nurtiti Sunusi, S.Si, M.Si
NIP. 19720117 199703 2002

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama	: Siti Hajriah
NIM	: H12116009
Program Studi	: Statistika
Jenjang	: S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul :

Pendugaan Area Kecil dengan Pendekatan Bayesian Hierarki

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 26 Januari 2021

Yang Menyatakan

METERAI TEMPEL

DB270AHF869185844

6000

ENAM RIBU RUPIAH

Siti Hajriah

PENDUGAAN AREA KECIL DENGAN PENDEKATAN BAYESIAN HIERARKI

Disetujui oleh:

Pembimbing Utama	**Pembimbing Pertama**
Dr.Dr. Georgina Maria Tinungki, M.Si	**Anisa, S.Si, M.Si**
NIP. 19620926 198702 2001	**NIP. 19730227 199802 2001**

Ketua Departemen Statistika,

Dr. Nurtiti Sunusi, S.Si, M.Si
NIP. 19720117 199703 2002

Pada Tanggal : 8 Desember 2020

KATA PENGANTAR

Alhamdulillahirobbil'alamin. Segala puji penulis haturkan atas kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul "Pendugaan Area Kecil dengan Pendekatan Bayesian Hierarki" ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains (S.Si) pada Pogram Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Univesitas Hasanuddin. Shalawat serta salam semoga tetap tercurahkan kepada junjungan kita Nabi Muhammad SAW karena Beliaulah yang membawa umat manusia dari kegelapan menuju ke alam yang terang benderang.

Penulis menyadari bahwa dalam penulisan skripsi ini tidak terlepas dari dorongan, dukungan, bimbingan, serta kerjasama dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penulisan skripsi ini. Terkhusus kepada orang tua **Ayahanda Samsudin** dan **Ibunda Rusia**, terimakasih telah membesarkan dan mendidik penulis dengan penuh kesabaran, memberikan limpahan cinta dan kasih sayang tiada tara, serta selalu memberi dukungan dan doa kepada penulis dalam menyelesaikan pedidikan. Kepada kakakku **Sarniati S** serta kedua adikku **Fifi Eviani** dan **Marsya** yang selalu membantu dan mendukung penulis. Serta untuk keluarga besar penulis, terima kasih pula atas doa dan dukungannya selama ini.

Demikian pula dengan penuh keikhlasan dan kerendahan hati penulis mengucapkan penghargaan dan terima kasih kepada :

1. **Ibu Prof. Dr. Dwia Aries Tina Palubuhu, MA,** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, S.Si, M.Si**, selaku Ketua Departemen Statistika serta segenap dosen pengajar dan staf Departemen Statistika yang telah membekali ilmu dan memberikan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Dr. Dr. Georgina Maria Tinungki, M.Si**, selaku Pembimbing Utama dan **Ibu Anisa, S.Si, M.Si** selaku Pembimbing Pertama sekaligus Penasehat

Akademik yang telah sabar dan ikhlas meluangkan begitu banyak waktu dan pemikirannya untuk membimbing dan memberikan masukan dalam penulisan skripsi ini.

5. **Ibu Sri Astuti Thamrin, S.Si, M.Stat, Ph.D** dan **Bapak Dr. La Podje Talangko, M.Si** selaku tim penguji atas semua saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada penulis.
6. Spesial untuk sahabat terdekat penulis yaitu **Mayang, Inayah, Devi** dan **Asma** yang telah menjadi teman seperantauan sekaligus keluarga selama berkuliah di Universitas Hasanuddin. Terima kasih telah mewarnai hari-hari penulis.
7. Teman-teman **Baubau-UH 2016** dan **Sciendrome** yang telah memberikan banyak dukungan dan motivasi kepada penulis.
8. **Donatur Paguyuban Karya Salemba Empat (KSE)** yang telah memberikan bantuan dana beasiswa yang tentunya sangat berguna bagi penulis. Terimakasih pula kepada **teman-teman Paguyuban KSE UNHAS** atas ilmu, kerja sama, dan dukungannya kepada penulis. Penulis sangat bersyukur karena telah dipertemukan dengan orang-orang hebat seperti kalian.
9. Sahabat seperjuangan **Statistika 2016** terkhusus **Tim Marketing (Iis, Wilda,** dan **Risma), Imma, Isna, Tari, Mila, Ririn, Jumri, Aco, Naim, Alam, Ibnu,** dan **Owi** yang banyak membantu penulis selama perkuliahan. Untuk **Tim Marketing**, terimakasih pula selalu bersedia menjadi tempat curhat penulis terlebih dalam proses penyelesaian skripsi ini.
10. Teman-teman **KKN Tematik UNHAS Kopi-Bulukumba Gel. 102** terkhusus posko Borrap. Terima kasih telah menjadi teman sekaligus keluarga selama sebulan lebih, semoga silaturahmi kita tetap terjalin.
11. Sepupuku **Sasmira Arif Masfarah** dan **Kakak Rurun** yang juga selalu membantu penulis selama berkuliah.
12. Terimakasih kepada **Kakak Tutun** yang sangat berbaik hati kepada penulis, selalu bersedia untuk mendengar curahan hati penulis dari waktu maba hingga saat ini.

13. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah Subhanahu Wata'ala.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini. Untuk itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Makassar, 8 Desember 2020

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: Siti Hajriah
NIM	: H 121 16 009
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Non-eksklusif (*Non-exclusiive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

"Pendugaan Area Kecil dengan Pendekatan Bayesian Hierarki"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 8 Desember 2020

Yang menyatakan,

Siti Hajriah

ABSTRAK

Pendugaan langsung *(direct estimation)* suatu survei seringkali tidak dapat diandalkan jika ukuran sampel yang tersedia terbatas sehingga data statistik yang dihasilkan juga tersedia pada level yang terbatas. Tujuan penelitian ini adalah penggunaan model Bayesian Hierarki yang diaplikasikan pada data pengangguran untuk menduga Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019 dengan model berbasis level area. Hasil yang diperoleh menunjukkan bahwa persentasi penduduk bekerja di sektor pertanian berpengaruh signifikan terhadap Tingkat Pengangguran Terbuka (TPT) di provinsi Sulawesi Selatan sehingga diperoleh model $log(\theta_i) = -4{,}27 - 0{,}01302x_2 + v_i$. Nilai estimasi Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan tahun 2019 sebesar 0,96 persen dimana Kota Palopo sebagai daerah dengan nilai estimasi tertinggi dan Kepulauan Selayar sebagai daerah dengan nilai estimasi terendah.

Kata Kunci : Bayesian Hierarki, Pendugaan Area Kecil, Tingkat Pengangguran Terbuka (TPT)

ABSTRACT

Direct estimation of a survey is often unreliable if the sample size available is limited so the resulting statistical data is also available at a limited level. The purpose of this research is the use of Hierarchical Bayesian model applied to unemployment data to suspect the open unemployment rate at the district/city level in South Sulawesi Province in 2019 with an area-level based model. The results showed that the percentage of people working in the agricultural sector has a significant effect on the open unemployment rate in South Sulawesi Province so that the model $log(\theta_i) = -4{,}27 - 0{,}01302x_2 + v_i$. The estimated value of open unemployment rate in South Sulawesi Province in 2019 is 0,96 percent where Palopo City as the region with the highest estimation value and Kepulauan Selayar as the region with the lowest estimated value.

Keywords: Hierarchical Bayesian, Open Unemployment Rate, Small Area Estimation

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang

Berbagai survei dilakukan guna membantu menyediakan data statistik yang cepat, lengkap, dan berkesinambungan. Beberapa alasan sehingga survei lebih banyak dipilih dibanding sensus adalah biaya yang digunakan lebih kecil, membutuhkan tenaga dan waktu yang jauh lebih sedikit, serta variabel atau karakteristik yang tercakup lebih banyak dan terinci. Namun, keterbatasan sampel pada data survei menyebabkan data statistik yang dihasilkan hanya tersedia pada level yang terbatas. Padahal banyak dari pengguna data membutuhkan data statistik yang memiliki tingkat pembahasan yang lebih rinci atau level yang lebih rendah (Ayuningtyas, 2017).

Menurut Rao (2003), pendugaan survei secara langsung *(direct estimation)* suatu indikator dengan jumlah sampel yang sedikit dalam suatu area kecil akan menyebabkan *standard error* yang besar. Pendugaan parameter dapat dilakukan sampai level yang lebih rendah jika sampel dari suatu survei dilakukan penambahan. Namun, banyaknya biaya yang dibutuhkan biasanya menjadi kendala sehingga penambahan sampel tidak dapat dilakukan. Oleh karena itu, diperlukan suatu metode tidak langsung yang dapat mengatasi hal tersebut (Ayuningtyas, 2017).

Pendugaan area kecil (*small area estimation*) merupakan suatu teknik statistik yang bertujuan untuk mendapatkan estimasi area kecil (atau domain) ketika pendugaan survei langsung (yaitu hanya berdasarkan pada data sampel spesifik area) tidak dapat diandalkan, (beberapa kali bahkan tidak dapat dihitung) karena ukuran sampel yang tersedia terbatas (Trevisani & Torelli, 2007). Area kecil yang dimaksud dapat berupa area geografis, seperti kabupaten/kota, kecamatan, kelurahan/desa, dan lain sebagainya. Area kecil juga dapat berupa kelompok sosial-demografi seperti jenis kelamin, kelompok ras, umur, dan sebagainya (Ayuningtyas, 2017). Pendugaan area kecil mempelajari dugaan tidak langsung dengan meminjam kekuatan *(borrow strength)* di seluruh area terkait, biasanya dengan sebuah model yang menghubungkan area-area yang ada melalui pemanfaatan informasi/variabel tambahan dari sensus atau survei lain yang berskala nasional (Rao, 2003).

Berbagai metode pendugaan area kecil telah banyak dikembangkan dan digunakan di berbagai bidang. Salah satu metode yang telah dikenal luas dan sering digunakan adalah Bayesian Hierarki. Bayesian Hierarki merupakan metode pendugaan area kecil yang berbasis model, dapat digunakan apabila variabel respon yang diteliti merupakan data kategorik (cacahan atau biner), dan memiliki keunggulan dibanding metode yang lain yaitu dapat mengatasi permasalahan pemodelan hirarki. Secara keseluruhan proses pemodelannya terbilang sangat rumit. Namun, setiap tahapnya relatif sederhana sehingga mudah dipahami (Ayuningtyas, 2017). Selain itu, Bayesian Hierarki juga dianggap sebagai alternatif yang baik dan diusulkan penggunaannya karena sangat sederhana yaitu secara komputasi jauh lebih efisien dengan menggunakan teknik *Markov Chain Monte Carlo* (MCMC) (Molina et al., 2014).

Berdasarkan penjelasan tersebut, maka dalam penelitian ini penulis akan mengkaji bagaimana memperoleh dugaan parameter dalam suatu area kecil dengan menggunakan Bayesian Hierarki. Pemodelan Bayesian Hierarki dalam penelitian ini akan diaplikasikan pada data pengangguran untuk menduga Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan. Menganggur adalah kejadian biner dengan dua kemungkinan, yaitu menganggur dan tidak menganggur sehingga Bayesian Hierarki dapat digunakan untuk menangani data tersebut.

Penelitian sebelumnya yang menggunakan metode pendugaan area kecil untuk menduga angka pengangguran dengan pendekatan hirarki diantaranya dilakukan oleh Juliyanto (2016). Dalam penelitiannya diterapkan beberapa model Bayesian Hierarki untuk menduga proporsi pengangguran tingkat kecamatan pada desain survei kompleks dengan distribusi variabel respon yang digunakan adalah distribusi Normal. Selain itu, Ayuningtyas (2017) melakukan penelitian untuk menduga proporsi pengangguran menurut kategori pengangguran di pulau Kalimantan menggunakan Bayesian Hierarki dengan variabel respon yang digunakan berdistribusi Multinomial dan Binomial.

Berdasarkan uji kesesuaian distribusi yang dilakukan terhadap data Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan tahun 2019 menggunakan uji Kolmogorov-Smirnov, diperoleh nilai statistik Kolmogorov-Smirnov untuk distribusi Log-Normal yang artinya data tersebut mengikuti distribusi Log-Normal. Penelitian yang dilakukan oleh Mauliani et al. (2018) juga menyebutkan bahwa data Tingkat Pengangguran Terbuka (TPT) yang digunakan mengikuti pola distribusi Log-Normal. Oleh karena itu, dalam penelitian ini penulis akan mengkaji bagaimana menduga parameter dengan pendekatan Bayesian Hierarki pada data pengangguran dimana variabel respon yang digunakan adalah berdistribusi Log-Normal.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka permasalahan yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana memperoleh model dugaan parameter dengan pendekatan Bayesian Hierarki untuk menduga Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan ?
2. Bagaimana hasil dugaan Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan dengan pendekatan Bayesian Hierarki ?

1.3 Batasan Masalah

Pada penelitian ini, permasalahan dibatasi pada :

1. Distribusi variabel respon yang digunakan adalah distribusi Log-Normal.
2. Model pendugaan area kecil yang digunakan merupakan model berbasis level area.

1.4 Tujuan Penulisan

Berdasarkan rumusan masalah, maka tujuan yang akan dicapai dalam penelitian ini adalah :

1. Memperoleh model dugaan parameter dengan pendekatan Bayesian Hierarki untuk menduga Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan.
2. Memperoleh hasil dugaan Tingkat Pengangguran Terbuka (TPT) level kabupaten/kota di Provinsi Sulawesi Selatan dengan pendekatan Bayesian Hierarki.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang hendak dicapai, maka penelitian ini diharapkan dapat bermanfaat bagi pembaca yaitu memberikan informasi mengenai pendugaan parameter dengan pendekatan Bayesian Hierarki khususnya pada kasus respon yang berdistribusi Log-Normal dan sebagai referensi pada penelitian-penelitian selanjutnya yang berhubungan dengan metode Bayes dalam menduga parameter untuk suatu area kecil.

BAB II
TINJAUAN PUSTAKA

2.1 Pendugaan Area Kecil

Pendugaan area kecil (*small area estimation*) merupakan suatu teknik statistik yang bertujuan untuk mendapatkan estimasi area kecil (atau domain) ketika pendugaan survei langsung (yaitu hanya berdasarkan pada data sampel spesifik area) tidak dapat diandalkan, (beberapa kali bahkan tidak dapat dihitung) karena ukuran sampel yang tersedia terbatas (Trevisani & Torelli, 2007). Suatu area atau domain disebut kecil apabila ukuran sampel dalam suatu area tidak cukup besar untuk mendukung pendugaan langsung dengan presisi yang memadai (Rao, 2003).

Berbagai survei umumnya dirancang untuk menduga parameter populasi untuk area yang besar, seperti level nasional atau provinsi dimana penduga parameternya didasarkan pada desain sampling. Hal ini menyebabkan umumnya jumlah sampel kurang/tidak mencukupi untuk menghasilkan pendugaan langsung yang akurat untuk pendugaan pada area kecil. Pendugaan langsung berarti melakukan pendugaan parameter populasi di suatu area hanya berdasarkan data yang diperoleh dari area tersebut (Rao, 2003). Karena pendugaan langsung tidak mampu memberikan ketelitian yang cukup bila ukuran sampel dalam area kecil yang menjadi perhatian sedikit, maka statistik yang dihasilkan akan memiliki varian yang besar atau bahkan pendugaan tidak dapat dilakukan karena tidak terwakili dalam survei (Prasad & Rao, 1990).

Untuk mengatasi masalah tersebut, maka diperlukan penggunaan data tambahan (seperti data sensus atau data yang berasal dari literatur atau penelitian sebelumnya) untuk mendapatkan dugaan yang akurat atau dapat dipercaya melalui suatu model tertentu. Pendugaan seperti ini disebut juga pendugaan tidak langsung *(indirect estimation)*, dalam arti bahwa pendugaan tersebut mencakup data dari area yang lain. Terdapat dua ide utama yang digunakan untuk mengembangkan model pendugaan parameter area kecil yaitu:

1. Asumsi bahwa keragaman di dalam area kecil variabel respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada

informasi tambahan, yang kemudian disebut model pengaruh tetap (*fixed effect models*).

2. Asumsi keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan dan merupakan pengaruh acak area kecil (*random effect*).

Gabungan dari dua asumsi tersebut membentuk model pengaruh campuran (*mixed models*) (Rao, 2003). Pendugaan area kecil memiliki dua jenis model dasar yaitu model berbasis level area dasar (*basic area level model*) dan model berbasis level unit dasar (*basic unit level model*) (Rao, 2003). Dalam penelitian ini, model yang digunakan adalah model berbasis level area dasar karena data respon yang digunakan merupakan data pada area tertentu yaitu level area kabupaten/kota.

2.2 Model Berbasis Level Area

Model berbasis level area merupakan model yang dapat digunakan bila data pendukung yang tersedia hanya ada untuk level area tertentu atau data pendukung yang bersesuaian dengan data yang diamati tidak tersedia hingga level unit contoh (R. Diana & Rory, 2020). Diasumsikan bahwa θ_i merupakan sebuah fungsi yang menghubungkan parameter $\boldsymbol{\beta}$ dengan variabel prediktor dari area kecil yaitu $\boldsymbol{x}_i = \left(x_{1i}, x_{2i}, \cdots, x_{pi}\right)^T$ yang merupakan suatu vektor. Model linier yang menjelaskan hubungan tersebut adalah (Rao, 2003) :

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i v_i \qquad (2.2.1)$$

dimana

$\boldsymbol{\beta}$: $\left(\beta_1, \beta_2, \cdots, \beta_p\right)^T$ merupakan parameter berukuran $p \times 1$ untuk data pendukung $\boldsymbol{x_i}$

i : area, $i = 1,2, \cdots, n$

n : banyaknya area

b_i : konstanta positif

v_i : pengaruh acak area ke-i yang diasumsikan memiliki distribusi $v_i \sim N(0, \sigma_v^2)$.

Jika $\widehat{\theta_i}$ adalah penduga langsung yang ada dalam model $\widehat{\theta_i} = \theta_i + e_i$, maka model (2.2.1) dapat ditulis menjadi :

$$\widehat{\theta_i} = \boldsymbol{x}_i^T\boldsymbol{\beta} + b_i v_i + e_i \tag{2.2.2}$$

dengan e_i adalah $sampling\ error$ pada area ke-i yang diasumsikan $e_i \sim N(0, \sigma_\varepsilon^2)$ dan σ_ε^2 diketahui.

Model pada persamaan (2.2.2) dikenal dengan model Fay-Herriot. Dalam model tersebut, keragaman variabel respon diasumsikan dapat diterangkan oleh hubungan variabel respon dengan informasi tambahan yang disebut sebagai model pengaruh tetap. Selain itu, terdapat komponen keragaman spesifik area kecil yang disebut sebagai komponen pengaruh acak area kecil. Gabungan dari dua asumsi tersebut membentuk model pengaruh campuran (Noviani, 2016).

2.3 Metode Bayes

Metode Bayes didasarkan pada dua sumber informasi dalam melakukan pendugaan parameter untuk suatu model statistik. Sumber informasi pertama berasal dari data sampel yang diperoleh dari suatu fungsi *likelihood* dan sumber informasi kedua berasal dari opini subjektif ahli atau berasal dari penelitian terdahulu yang disebut dengan informasi *prior*. Berbagai ahli dapat memberikan informasi *prior* berbeda. Ketidaktentuan opini ahli ini diekspresikan melalui suatu distribusi peluang yang disebut distribusi *prior* (Bolstad, 2007). Pada metode ini untuk menduga parameter dilakukan pendekatan dengan memperlakukan semua parameter yang tidak diketahui sebagai peubah acak dan memiliki distribusi (Bolstad, 2007). Dari distribusi *posterior* diperoleh estimator Bayes yang merupakan *mean* dari distribusi *posterior*. Hubungan distribusi *posterior* dengan distribusi *prior* dan fungsi *likelihood* yaitu (Ntzoufras, 2009) :

$$\text{Distribusi}\, posterior \propto likelihood \times \text{distribusi}\, prior$$

Notasi $\propto$ menyatakan suatu hubungan proposionalitas dari distribusi *posterior.*

2.4 Distribusi *Posterior*

Distribusi *posterior* adalah fungsi kepadatan peluang bersyarat dari θ jika diketahui nilai observasinya, dimisalkan $\boldsymbol{y} = y_1, y_2, \cdots, y_n$ dapat ditulis sebagai :

$$f(\theta|\boldsymbol{y}) = \frac{f(\theta,\boldsymbol{y})}{f(\boldsymbol{y})} \qquad (2.4.1)$$

Fungsi kepadatan peluang bersama, $f(\theta, \boldsymbol{y})$ dapat ditulis dalam bentuk :

$$f(\theta, \boldsymbol{y}) = f(\boldsymbol{y}|\theta)\, f(\theta)$$

dimana $f(\boldsymbol{y}|\theta)$ adalah fungsi *likelihood* dan $f(\theta)$ merupakan distribusi *prior* bagi θ. Selanjutnya, fungsi kepadatan peluang marginal bagi $\boldsymbol{y}$ dapat dinyatakan sebagai

$$f(\boldsymbol{y}) = \int_{-\infty}^{\infty} f(\theta, \boldsymbol{y})\, d\theta = \int_{-\infty}^{\infty} f(\boldsymbol{y}|\theta)\, f(\theta)\, d\theta \qquad (2.4.2)$$

Dari persamaan (2.5.1) dan (2.5.2), fungsi kepadatan peluang *posterior* dapat didefinisikan sebagai berikut.

Definisi 2.5.1 : (Bain & Engelhardt, 1992)

Fungsi kepadatan peluang bersyarat dari θ jika diketahui sampel observasi $y_1, y_2, \cdots, y_n$ disebut sebagai fkp *posterior* yang didefinisikan sebagai berikut.

$$f(\theta|\boldsymbol{y}) = \frac{f(\theta,\boldsymbol{y})}{\int f(\theta,\boldsymbol{y})\, d\theta} \propto f(y_1, y_2, \cdots, y_n|\theta)\, f(\theta) \qquad (2.4.3)$$

2.5 Fungsi *Likelihood*

Definisi 2.6.1 (Bain & Engelhardt, 1992)

Fungsi *likelihood* adalah fungsi kepadatan peluang bersama dari n peubah acak $Y_1, Y_2, \cdots, Y_n$ yang dihitung pada $y_1, y_2, \cdots, y_n$ dan dinyatakan dalam bentuk $f(y_1, y_2, \cdots, y_n|\theta)$. Jika $y_1, y_2, \cdots, y_n$ ditetapkan, maka fungsi *likelihood* adalah fungsi dari parameter θ dan dinotasikan dengan $L(\theta)$. Jika $Y_1, Y_2, \cdots, Y_n$ menyatakan suatu peubah acak dari $f(y;\theta)$, maka:

$$L(\theta) = f(y_1;\theta)\, f(y_2;\theta) \cdots f(y_n;\theta)$$
$$= \prod_{i=1}^{n} f(y_i;\theta) \qquad (2.5.1)$$

2.6 Distribusi *Prior*

Distribusi *prior* merupakan distribusi awal yang harus diketahui untuk menentukan distribusi *posterior* suatu data. Permasalahan utama dalam metode Bayes adalah bagaimana memilih distribusi *prior* dimana *prior* menunjukkan ketidakpastian tentang parameter yang tidak diketahui. Ada beberapa tipe distribusi *prior* dalam metode Bayes, yaitu sebagai berikut (Box & Tiao, 1973) :

1. Berdasarkan bentuk identifikasi pola data yang diperoleh dari fungsi *likelihood* nya, terdiri dari :
   a. Distribusi *prior* konjugat

      Distribusi *prior* ini mengacu pada acuan analisis model terutama dalam pembentukan fungsi *likelihood* nya sehingga dalam penentuan *prior* konjugat selalu dipikirkan mengenai penentuan pola distribusi *prior* yang mempunyai bentuk konjugat dengan fungsi kepadatan peluang pembangun *likelihood* nya. Beberapa bentuk fungsi kepadatan peluang (fkp) distribusi *prior* konjugat dapat dilihat pada tabel berikut (Ntzoufras, 2009).

Tabel 2.6.1 Bentuk Fungsi Kepadatan Peluang Konjugat untuk Beberapa Fungsi Kepadatan Peluang

No.	Fkp	Fkp Konjugat
1.	Binomial (p,N)	p ~ beta(a, b)
2.	Poisson (λ)	λ ~ gamma(a, b)
3.	N (θ, σ^2), σ^2 known	θ ~ N(θ_0, σ_0^2)
4.	N (θ, σ^2), θ known, $\tau = {}^{1}\!/_{\sigma^2}$	τ ~ gamma(a, b)
5.	Gamma (v, θ), v known	θ ~ gamma(a, b)
6.	Beta (v, θ), θ known	v ~ gamma(a, b)

Sumber : Ntzoufras (2009)

   b. Distribusi *prior non*-konjugat

      Distribusi *prior non*-konjugat tidak mengindahkan pola pembentukan fungsi *likelihood* nya (Ntzoufras, 2009).

2. Berdasarkan informasi terdahulu terkait dengan penentuan masing-masing parameter pada pola distribusi *prior* nya, yaitu:
   a. Distribusi *prior* informatif

      Distribusi *prior* informatif mengacu pada pemberian parameter dari distribusi *prior* yang telah dipilih baik distribusi *prior* konjugat atau tidak. Pemberian nilai parameter pada distribusi *prior* ini akan sangat mempengaruhi bentuk distribusi *posterior* yang akan didapatkan pada informasi data yang diperoleh. Distribusi *prior* informatif ditentukan berdasarkan pada ketersediaan informasi sebelumnya (penelitian terdahulu) mengenai pola distribusi data. (Ntzoufras, 2009).

b. Distribusi *prior non*-informatif

Distribusi *prior non*-informatif digunakan apabila pemilihan distribusi *prior* nya tidak didasarkan pada data yang ada sebelumnya atau distribusi *prior* yang tidak mengandung informasi tentang parameter.

2.7 Distribusi Normal

Definisi 2.8.1 : (Bain & Engelhardt, 1992)

Suatu peubah acak X dikatakan berdistribusi Normal dengan *mean* θ dan varians σ^2 jika memiliki fungsi kepadatan peluang dalam bentuk :

$$f(x|\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right) \qquad (2.7.1)$$

untuk $-\infty < x < \infty, -\infty < \theta < \infty$, dan $0 < \sigma < \infty$.

Peubah acak yang berdistribusi Normal dengan parameter θ dan σ^2 dapat dinyatakan dengan $X \sim N(\theta, \sigma^2)$.

2.8 Distribusi Log-Normal

Distribusi Log-Normal adalah distribusi kontinu positif. Distribusi Log-Normal merupakan distribusi tidak simetri/miring yang mana probabilitas data-data lainnya tersebar tidak secara merata. Grafik fungsi kepadatannya tidak berbentuk lonceng, tidak seperti distribusi Normal (E. N. Diana, 2016). Distribusi Log-Normal diperoleh dari transformasi peubah acak pada fungsi kepadatan distribusi Normal.

Definisi 2.8.1 : (Crow & Shimizu, 1988)

Misalkan sebuah peubah acak Y bilangan real positif $(0 < y < \infty)$ sedemikian sehingga $X = \ln Y$ merupakan distribusi Normal dengan *mean* θ dan varians σ^2. $Y = e^X$ merupakan distribusi Log-Normal atau dapat ditulis dengan $LN(\theta, \sigma^2)$ dan $X \sim N(\theta, \sigma^2)$. Fungsi kepadatan peluang Log-Normal adalah sebagai berikut:

$$f(y|\theta,\sigma^2) = \begin{cases} \frac{1}{y\sigma\sqrt{2\pi}}\, exp\left(-\frac{1}{2\sigma^2}(\ln(y)-\theta)^2\right), & y > 0 \\ 0 & , \ y \le 0 \end{cases} \qquad (2.8.1)$$

Teorema 2.8.1 : (E. N. Diana, 2016)

Variabel acak positif Y yang berdistribusi Log-Normal mempunyai *mean* $exp\left(\theta+\frac{\sigma^2}{2}\right)$ dan varians $exp(2\theta+\sigma^2)(exp(\sigma^2)-1)$.

Bukti.

Perhatikan bahwa :

$$E(Y) = \int_{-\infty}^{\infty} y\, f(y)\, dy$$

$$= \int_{-\infty}^{0} y\, f(y)\, dy + \int_{0}^{\infty} y\, f(y)\, dy$$

Berdasarkan Definisi 2.8.1, jika $y \le 0$ maka $\int_{-\infty}^{0} y\, f(y)\, dy = 0$ sehingga

$$E(Y) = \int_{0}^{\infty} y\, f(y)\, dy$$

$$= \int_0^\infty y\, \frac{1}{y\sigma\sqrt{2\pi}}\, exp\left(-\frac{1}{2\sigma^2}(\ln(y)-\theta)^2\right) dy$$

$$= \int_0^\infty y\, \frac{1}{y\sqrt{2\pi\sigma^2}}\, exp\left(-\frac{1}{2}\left(\frac{\ln(y)-\theta}{\sigma}\right)^2\right) dy$$

Misalkan, $z = \left(\frac{\ln(y)-\theta}{\sigma}\right)$ maka $\ln(y) = z\sigma+\theta$ sehingga $y = exp(z\sigma+\theta)$ dan $dy = \sigma\, exp(z\sigma+\theta)dz$. Dengan transformasi tersebut, untuk $y \to 0^+$ maka $z \to -\infty$ dan untuk $y \to \infty$ maka $z \to \infty$, sehingga :

$$E(Y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, exp\left(-\frac{1}{2}z^2\right) \sigma\, exp(z\sigma+\theta)dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, exp\left(-\frac{z^2}{2}+z\sigma+\theta\right) dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, exp\left(-\frac{z^2}{2}+z\sigma+\theta+\frac{\sigma^2}{2}-\frac{\sigma^2}{2}\right) dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}(z-\sigma)^2+\left(\theta+\frac{\sigma^2}{2}\right)\right) dz$$

$$= exp\left(\theta+\frac{\sigma^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{2}(z-\sigma)^2\right) dz$$

Karena $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\left(-\frac{1}{2}(z-\sigma)^2\right) dz = 1$, sehingga diperoleh bahwa

$E(Y) = exp\left(\theta+\frac{\sigma^2}{2}\right)$.

Jadi, terbukti bahwa untuk $Y \sim LN(\theta, \sigma^2)$ maka *mean* $\theta = E(Y) = exp\left(\theta + \frac{\sigma^2}{2}\right)$

Selanjutnya, akan dicari varians dari distribusi Log-Normal :

$\sigma^2 = E(Y^2) - (E(Y)^2)$

$$\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 \, f(y) \, dy \\
&= \int_{-\infty}^{\infty} y^2 \, \frac{1}{y\sigma\sqrt{2\pi}} \, exp\left(-\frac{1}{2\sigma^2}(\ln(y) - \theta)^2\right) dy \\
&= 0 + \int_{0}^{\infty} y \frac{1}{\sigma\sqrt{2\pi}} \, exp\left(-\frac{1}{2\sigma^2}(\ln(y) - \theta)^2\right) dy \\
&= \int_{0}^{\infty} y \, \frac{1}{\sigma\sqrt{2\pi}} \, exp\left(-\frac{1}{2}\left(\frac{\ln(y) - \theta}{\sigma}\right)^2\right) dy
\end{aligned}$$

Misalkan, $z = \frac{\ln(y) - \theta}{\sigma}$ atau $\ln(y) = z\sigma + \theta$ maka $y = exp(z\sigma + \theta)$ dan $dy = \sigma \, exp(z\sigma + \theta) dz$. Dengan transformasi tersebut, untuk $y \to 0^+$ maka $z \to -\infty$ dan untuk $y \to \infty$ maka $z \to \infty$, sehingga

$$\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} \frac{exp(z\sigma + \theta)}{\sqrt{2\pi}\,\sigma} exp\left(-\frac{1}{2}z^2\right) \sigma \, exp(z\sigma + \theta) \, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2} + 2z\sigma + 2\theta\right) dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2} + 2z\sigma + 2\theta + 2\sigma^2 - 2\sigma^2\right) dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left(-\frac{1}{2}(z - 2\sigma)^2 + 2\theta + 2\sigma^2\right) dz \\
&= exp(2\theta + 2\sigma^2) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}(Z - 2\sigma)^2\right) dz \\
&= exp(2\theta + 2\sigma^2)
\end{aligned}$$

Sehingga, $\sigma^2 = Var(Y) = E(Y^2) - (E(Y)^2)$

$$\begin{aligned}
&= exp(2\theta + 2\sigma^2) - \left(exp\left(\theta + \frac{\sigma^2}{2}\right)\right)^2 \\
&= exp(2\theta + 2\sigma^2) - exp(2\theta + \sigma^2) \\
&= exp(2\theta + \sigma^2)(exp(\sigma^2) - 1)
\end{aligned}$$

Jadi, terbukti bahwa varians $= exp(2\theta + \sigma^2)(exp(\sigma^2) - 1)$

2.9 Bayesian Hierarki dengan Model Log-Normal - Normal

Bayesian Hierarki merupakan salah satu metode pendugaan area kecil yang paling sering digunakan. Metode ini mengasumsikan bahwa distribusi *prior* $f(\theta)$ dari parameter model diketahui. Disini juga data dari survei **y** harus disertakan. Dengan menggunakan teorema Bayes kita dapat memperoleh distribusi *posterior* $f(\theta|\boldsymbol{y})$. Dalam kasus sederhana distribusi tersebut dapat diperoleh secara analitik. Namun, kasus yang lebih kompleks memerlukan metode komputasi khusus dengan menggunakan teknik *Markov Chain Monte Carlo* (MCMC) yang diimplementasikan secara numerik menggunakan algoritma *Gibbs Sampling* (Kubacki, 2012).

Pendugaan parameter dengan Bayesian Hierarki memiliki beberapa keunggulan yaitu pemodelannya dilakukan secara bertahap. Secara keseluruhan proses pemodelannya terbilang sangat rumit, namun setiap tahapnya relatif sederhana sehingga mudah dipahami (Ayuningtyas, 2017). Bayesian Hierarki ini dapat mengatasi permasalahan pemodelan hirarki untuk jumlah data yang sedikit dan tidak setimbang, baik pada tingkat bawah maupun tingkat yang lebih tinggi.

Orme (2000) menyebutkan bahwa dibanding regresi berganda, penggunaan Bayesian Hierarki jauh lebih baik dalam menghasilkan dugaan parameter yang reliabel dan akurat. Metode ini juga dapat menggabungkan model yang kompleks dimana metode lain tidak dapat menanganinya seperti *unmatched sampling and linking model* dan model-model yang mengasumsikan pengaruh acak *(random effect)* mengikuti distribusi selain distribusi Normal (Raudenbush & Bryk, 2002).

Model yang digunakan dalam penelitian ini merupakan model berbasis level area. Diketahui $y_i = \widehat{\theta_i}$ adalah pendugaan langsung *(direct estimation)* dari parameter area ke-i. Dalam penelitian ini, y_i diasumsikan berdistribusi Log-Normal atau dapat ditulis $y_i \sim LN\left(\theta_i, \sigma^2_{[y]}\right)$. Li et al., (2020) dalam penelitiannya menggunakan model berbasis level area dengan data respon yang digunakan diasumsikan mengikuti distribusi Log-Normal 3 parameter (LN3) dan kemudian melakukan transformasi log sehingga data yang ditransformasi mengikuti distribusi Normal. Transformasi log pada y_i

dilakukan agar peubah Log-Normal dapat diubah menjadi peubah Normal yang merupakan properti yang diinginkan untuk membangun model hirarki pada Bayesian (Li et al., 2020). Adapun model yang digunakan pada penelitian ini adalah model Log-Normal - Normal yang dituliskan sebagai berikut.

1. *Sampling model*

   $\log(y_i) \sim N\left(\log(\theta_i), \sigma^2_{[y]}\right)$, dimana $\sigma^2_{[y]} = \frac{1}{\tau_{[y]}}$

2. *Linking model*

   $\log(\theta_i) \sim N\left(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2_{[v]}\right)$, dengan $\log(\theta_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i$, $v_i \sim N(0, \sigma^2_{[v]})$ dan $\sigma^2_{[v]} = \frac{1}{\tau_{[v]}}$

3. $\boldsymbol{\beta}$ dan $\sigma^2_{[v]}$ saling bebas dengan $\frac{1}{\sigma^2_{[v]}} \sim Gamma(a, b)$ dan $\beta_j \sim N(\mu_{[\beta]}, \sigma^2_{[\beta]})$

Model dalam penelitian ini termasuk ke dalam *matched model*, dalam arti bahwa antara *sampling* dan *linking model* dapat dikombinasikan untuk menghasilkan model linier campuran seperti pada model (2.2.2) (Trevisani & Torelli, 2007). Adapun yang dimaksud dengan konsep hirarki pada model Log-Normal - Normal ini adalah pengaruh acak *(random effect)* v_i yang diasumsikan mengikuti distribusi Normal yaitu $v_i \sim N(0, \sigma^2_{[v]})$ dan *hyperparameter* varians dari pengaruh acak diasumsikan mengikuti distribusi tertentu yang disebut dengan *hyperprior* (misal $\sigma^{-2}_{[v]} \sim Gamma(a, b)$).

2.10 *Markov Chain Monte Carlo* (MCMC)

Pada metode Bayes, pendugaan parameter diperoleh dengan menggunakan distribusi *posterior*. Proses penentuan distribusi *posterior* dari fungsi *likelihood* dan distribusi *prior* seringkali sulit dilakukan jika melibatkan persamaan integral yang sangat kompleks. Dalam pendekatan Bayesian, hal ini biasanya diatasi melalui penggunaan *Markov Chain Monte Carlo* (MCMC) (Carlin & Chib, 1995).

Ide dasar dari MCMC adalah membangkitkan data sampel dari distribusi *posterior* sesuai proses *Markov 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 tertentu sehingga tidak dapat digunakan untuk membangkitkan sampel dari berbagai bentuk distribusi *posterior* yang ada.

Misal $\boldsymbol{\beta}^{(t)}$ merupakan suatu nilai pada periode ke-t dan diinginkan suatu distribusi dari deret pada nilai-nilai berurutan $\boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \cdots$ serta diasumsikan bahwa $\boldsymbol{\beta}^{(t)}$ merupakan peubah acak yang saling independen. Misal $\{\boldsymbol{\beta}^{(t)}, t = 0,1,2, \cdots\}$ merupakan proses stokastik. Jika $\boldsymbol{\beta}^{(t)} = i$, maka proses dikatakan dalam keadaan i pada waktu ke-t . Dapat dituliskan seperti persamaan (2.11.1).

$$f\left(\boldsymbol{\beta}^{(t+1)} \middle| \boldsymbol{\beta}^{(t)}, \cdots, \boldsymbol{\beta}^{(1)}\right) = f\left(\boldsymbol{\beta}^{(t+1)} \middle| \boldsymbol{\beta}^{(t)}\right) \quad (2.10.1)$$

Proses stotatistik ini dikenal sebagai *Markov chain* (Ross, 2014). Persamaan (2.11.1) dapat diinterpretasikan bahwa distribusi bersyarat untuk semua keadaan selanjutnya $\boldsymbol{\beta}^{(t+1)}$ dengan syarat keadaan sebelumnya $\boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \cdots, \boldsymbol{\beta}^{(t-1)}$ dan keadaan saat ini $\boldsymbol{\beta}^{(t)}$ bersifat independen dengan keadaan sebelumnya dan hanya bergantung pada keadaan saat ini. Pada saat $t \to \infty$, distribusi dari $\boldsymbol{\beta}^{(t)}$ akan konvergen menuju distribusi tertentu yang independen terhadap nilai awal dari rantai tersebut, $\boldsymbol{\beta}^{(0)}$ (Ntzoufras, 2009).

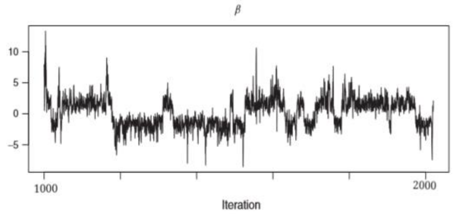
Proses MCMC dilakukan dengan cara membangkitkan *Markov Chain* yang konvergen terhadap distribusi target yaitu distribusi *prior* dari parameter yang diduga (disebut kondisi stasioner atau kondisi *equilibrium*). Sampel parameter dalam *Markov Chain* diambil setelah kondisi *equilibrium* tercapai, sehingga sampel yang diperoleh dijamin merupakan sampel dari distribusi target yaitu distribusi *prior* dari parameter tersebut. Kondisi *equilibrium* tercapai jika sampel yang diperoleh telah memenuhi sifat dari *Markov Chain* yang *strongly ergodic*, yaitu:

a. *Irreducible*, yaitu artinya sampel parameter yang dibangkitkan melalui proses MCMC bersifat acak
b. *Aperiodic*, yaitu artinya sampel parameter yang dibangkitkan tidak memiliki pola yang periodik dalam domain tertentu.

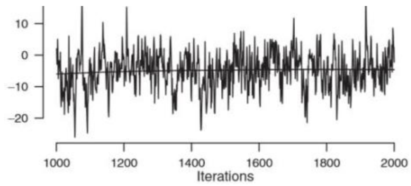
c. *Recurrent,* artinya perubahan sampel parameter terjadi secara stabil dalam domain nilai tertentu.

Ketiga sifat *Markov Chain* tersebut dapat dideteksi melalui *diagnostic plot* yang diperoleh dari proses MCMC (Ntzoufras, 2009), yaitu :

1. *Trace plot* adalah plot dari pergerakan nilai sampel yang diperoleh untuk sebagian iterasi saja, sedangkan pergerakan nilai sampel untuk seluruh iterasi diilustrasikan melalui *serial* plot. Jika semua nilai-nilai berada dalam sebuah daerah yang tidak terdapat pola periodik, maka sampel yang diperoleh telah memenuhi sifat MCMC yang *strongly ergodic*.

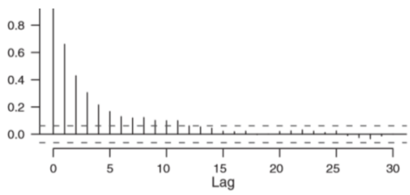


Gambar 2.10.1 Pola berperiodik dan tidak konvergen

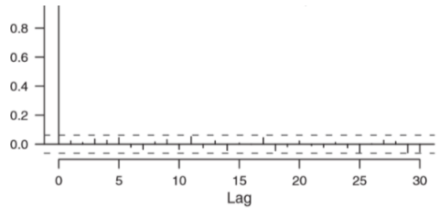


Gambar 2.10.2 Pola aperiodik

2. *Autocorrelation plot* adalah plot dari nilai *autocorrelation function* (ACF). Jika nilai ACF data sampel MCMC mendekati nol, maka sampel yang diperoleh telah memenuhi sifat *strongly ergodic*. Untuk contoh kekonvergenan *autocorrelation* plot dapat dilihat pada **Gambar 2.10.3** dan **Gambar 2.10.4**

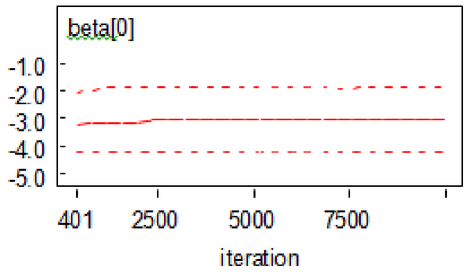


Gambar 2.10. 3 Pola kekonvergenan berjalan lambat



Gambar 2.10.4 Pola kekonvergenan berjalan cepat

3. *Quantiles plot* adalah plot dari *ergodic mean* dari parameter yang diduga. *Ergodic mean* adalah *mean* dari dugaan *posterior* selama proses iterasi berlangsung. Jika nilai *ergodic mean* parameter telah mencapai nilai yang stabil dan semua nilai sampel berada dalam *credible interval*, maka sampel yang telah diperoleh memenuhi sifat MCMC yang *strongly ergodic*.



Gambar 2.10.5 *Quantiles plot* menunjukkan data sampel stabil dan berada pada *credible interval*

2.11 *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 peubah acak dalam MCMC adalah *Gibbs Sampling*. *Gibbs sampling* dapat didefinisikan sebagai suatu teknik simulasi untuk membangkitkan peubah acak dari suatu fungsi distribusi tertentu tanpa harus menghitung fungsi densitasnya (Casella & George, 1992).

Gibbs sampler merupakan generator yang sangat efisien sehingga sering digunakan sebagai generator peubah acak pada analisis data yang menggunakan MCMC. Proses ini dilakukan dengan mengambil sampel dengan cara membangkitkan rangkaian *Gibbs* peubah acak 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 peubah acak 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 dan sulit diselesaikan secara analitik.

Jika $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \cdots, \beta_p\right)^T$ merupakan suatu vektor dari sekumpulan parameter yang akan diduga, maka dapat dibentuk distribusi *full conditional posterior* untuk setiap parameter dari $\boldsymbol{\beta}$. Distribusi *full conditional posterior* untuk β_j atau $f\left(\beta_j \middle| \boldsymbol{\beta}_{\backslash j}, \mathbf{y}\right)$ dibentuk dari distribusi *posterior* gabungan seluruh parameter pada $\boldsymbol{\beta}$, yaitu dengan menetapkan nilai parameter selain β_j atau $\boldsymbol{\beta}_{\backslash j} = \left(\beta_1, \beta_2, \cdots, \beta_{j-1}, \beta_{j+1}, \cdots, \beta_p\right)^T$ bernilai konstan dan $\boldsymbol{\beta}_{\backslash j}$ merupakan vektor $\boldsymbol{\beta}$ tanpa elemen β_j. Karena nilai parameter $\boldsymbol{\beta}_{\backslash j}$ dianggap konstan, maka $f\left(\beta_j \middle| \boldsymbol{\beta}_{\backslash j}, \mathbf{y}\right)$ adalah bentuk proposional dari distribusi *posterior* gabungan seluruh parameter yang hanya mengandung komponen parameter β_j dan komponen yang mengandung parameter selain β_j dieliminasi. Selanjutnya, distribusi *full conditional posterior* untuk parameter β_j inilah yang akan digunakan dalam proses iterasi pendugaan parameter β_j menggunakan MCMC dengan algoritma *gibbs sampling*. Berikut diberikan algoritma MCMC dan *gibbs sampling*, yaitu (Ntzoufras, 2009) :

1. Tetapkan nilai awal (*initial value*) parameter $\boldsymbol{\beta}^{(0)}$.
2. Tentukan banyaknya iterasi (M) untuk membangkitkan sampel masing-masing parameter yang akan diestimasi. Untuk $t = 1,2, \cdots, M$, ulangi langkah:
   a. Tentukan $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)}$.
   b. Untuk $j = 0, \cdots, d$, *update* β_j dari $\beta_j \sim f\ \left(\beta_j \middle| \boldsymbol{\beta}_{\backslash j}, \boldsymbol{y}\right)$
   c. Tentukan $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}$ dan gunakan untuk membangkitkan iterasi ke $t + 1$.

Berikut ini adalah proses *sampling* untuk mendapatkan nilai $\boldsymbol{\beta}^{(t)}$ dari distribusi *full conditional posterior,* yaitu:

Bangkitkan $\beta_0^{(t)}$dari $f\left(\beta_0\middle|\beta_1^{(t-1)},\beta_2^{(t-1)},\cdots,\beta_p^{(t-1)},\boldsymbol{y}\right)$,

Bangkitkan $\beta_1^{(t)}$dari $f\left(\beta_1\middle|\beta_2^{(t-1)},\beta_3^{(t-1)},\cdots,\beta_p^{(t-1)},\boldsymbol{y}\right)$,

Bangkitkan $\beta_2^{(t)}$dari $f\left(\beta_2\middle|\beta_1^{(t)},\beta_3^{(t-1)},\cdots,\beta_p^{(t-1)},\boldsymbol{y}\right)$,

Bangkitkan $\beta_3^{(t)}$dari $f\left(\beta_3\middle|\beta_1^{(t)},\beta_2^{(t)},\beta_4^{(t-1)},\cdots,\beta_p^{(t-1)},\boldsymbol{y}\right)$,

⋮

Bangkitkan $\beta_j^{(t)}$dari $f\left(\beta_j\middle|\beta_1^{(t)},\beta_2^{(t)},\cdots,\beta_{j-1}^{(t)},\beta_{j+1}^{(t-1)},\cdots,\beta_p^{(t-1)},\boldsymbol{y}\right)$,

⋮

Bangkitkan $\beta_p^{(t)}$dari $f\left(\beta_p\middle|\beta_0^{(t)},\beta_1^{(t)},\cdots,\beta_{p-1}^{(t)},\boldsymbol{y}\right)$.

3. Amati konvergenitas data sampel. Jika kondisi konvergen belum tercapai, maka dibutuhkan sampel lebih banyak lagi dan ulangi langkah 2 sampai kondisi konvergen
4. Melakukan proses *burn in* dengan membuang sebanyak C sampel pertama. Proses *burn in* adalah periode awal iterasi estimasi parameter dalam proses MCMC membuang sebanyak C iterasi pertama, guna menghilangkan pengaruh dari penggunaan nilai awal. Periode *burn in* dilakukan sampai kondisi stasioner tercapai. Dari proses MCMC akan diperoleh sampel sebanyak $D = M - C$ untuk setiap parameter.
5. Menggunakan $\left\{\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)},\cdots\boldsymbol{\beta}^{(D)}\right\}$ sebagai sampel dalam analisis *posterior.*
6. Membuat plot distribusi *posterior*.
7. Membuat *summary* dari distribusi *posterior* (*mean*,standar deviasi, dll).

2.12 Evaluasi Penduga

Penduga yang baik adalah penduga yang memiliki nilai *standard error* yang kecil. *Standard error* dapat menunjukkan seberapa akurat dalam menduga parameter yang dirumuskan sebagai berikut (Lee et al., 2015) :

$$SE(\hat{\theta}) = \frac{SD(\hat{\theta})}{\sqrt{n}}; SD(\hat{\theta}) = \sqrt{Var\,\hat{\theta}} \qquad (2.12.1)$$

$SD(\hat{\theta})$ adalah standar deviasi dari $\hat{\theta}$, n adalah banyaknya sampel, dan $Var\ \hat{\theta}$ adalah ragam penduga.

2.13 Uji Kolmogorov-Smirnov

Tahapan awal yang dilakukan dalam penelitian ini adalah melakukan uji kecocokan distribusi (*goodness of fit*) untuk mengetahui distribusi data dari variabel respon. Uji kecocokan distribusi terdiri dari tiga jenis yaitu uji Bartlett untuk distribusi Eksponensial, uji Mann untuk distribusi Weibull, dan uji Kolmogorov-Smirnov untuk distribusi Normal dan Log-Normal (Famela et al., 2017).

Dalam penelitian ini uji kecocokan distribusi yang digunakan adalah uji Kolmogrov-Smirnov. Uji Kolmogrov-Smirnov adalah uji statistik yang digunakan untuk menguji tingkat kesesuaian atau ketidaksesuaian antara distribusi pada data sampel (data yang diobservasi) terhadap suatu distribusi teoritis tertentu. Misalkan $F(y)$ adalah suatu fungsi distribusi (cdf) yang sepenuhnya ditentukan secara teoritis, dan $G(y)$ adalah suatu distribusi yang diobservasi berdasarkan suatu data sampel.

Berdasarkan hipotesis nol (H_0) bahwa sampel itu telah disimpulkan berasal dari distribusi teoritis tertentu, maka diharapkan bahwa untuk setiap $G(y)$ harus mendekati $F(y)$. Artinya, berdasarkan H_0 akan diharapkan selisih antara $G(y)$ dan $F(y)$ adalah kecil dan berada pada batas-batas kesalahan acak. Uji Kolmogrov-Smirnov ini menggunakan statistik uji yaitu pada penyimpangan terbesar atau disebut deviasi maksimum antara $G(y)$ dan $F(y)$. Misalkan deviasi maksimum tersebut adalah S_n , maka :

$$S_n = sup|F(y) - G(y)| \tag{2.13.1}$$

Hipotesis yang digunakan adalah

$$H_0: F(y) = G(y)$$

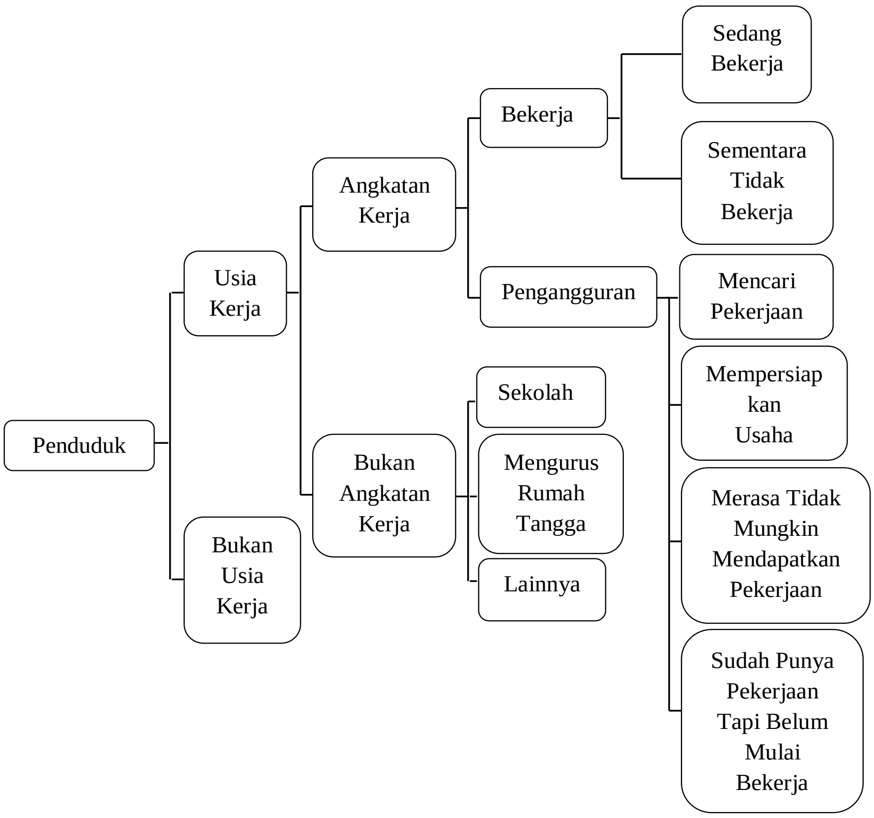
$$H_1: F(y) \neq G(y)$$

Kriteria pengambilan keputusannya adalah H_0 akan ditolak jika $S_n > S_{n,\alpha}$ dengan $S_{n,\alpha}$ adalah nilai kritik yang diperoleh dari tabel Kolmogrov-Smirnov atau H_0 akan ditolak jika *p-value* $< \alpha$. (Lehmann & Romano, 2005)

2.14 Pengangguran Terbuka

2.14.1 Konsep Pengangguran

Pengangguran merupakan suatu keadaan dimana seseorang yang termasuk dalam kategori angkatan kerja tidak memiliki pekerjaan dan secara aktif sedang mencari pekerjaan (Nanga, 2001). Menurut Pitartono & Hayati (2012), pengangguran dapat terjadi sebagai akibat dari persentasi penyerapan tenaga kerja yang cenderung kecil serta tingkat perubahan angkatan kerja yang tinggi tidak diimbangi dengan pertumbuhan penciptaan lapangan kerja.



Gambar 2.14.1 Konsep Pengangguran (Badan Pusat Statistik, 2018)

Tingkat Pengangguran Terbuka (TPT) adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja.

$$\text{TPT} = \frac{Jumlah\ pengangguran}{jumlah\ angkatan\ kerja} \times 100\% \qquad (2.14.1)$$

Tingkat Pengangguran Terbuka (TPT) atau biasa disebut dengan tingkat pengangguran menggambarkan proporsi angkatan kerja yang tidak memiliki pekerjaan dan secara aktif mencari dan bersedia untuk bekerja (Badan Pusat Statistik, 2018)

2.14.2 Variabel Prediktor yang Mempengaruhi Pengangguran

Rao (2003) mengatakan bahwa penggunaan variabel prediktor yang berpengaruh terhadap data yang diobservasi memiliki peran penting untuk menghasilkan dugaan yang lebih akurat dalam pendugaan tidak langsung suatu area kecil. Kriteria penggunaan variabel prediktor ini berdasarkan pada literatur, ketersediaan data, dan penelitian tentang pengangguran serta ketenagakerjaan yang pernah dilakukan sebelumnya.

Menurut Rompas et al. (2015), sektor pertanian merupakan salah satu sektor yang memiliki potensi dalam kegiatan perekonomian. Pemerintah diharapkan mampu mengembangkan potensi sektor pertanian sehingga dapat menciptakan lapangan kerja yang akan berpengaruh terhadap penyerapan tenaga kerja serta membantu perekonomian daerah. Rini (2012) menyatakan bahwa sektor informal memiliki peran aktif dalam meningkatkan keterserapan jumlah pengangguran. Sektor ini bahkan memberikan nilai positif dalam angka statistik aspek ketenagakerjaan.

Edy (2009) melakukan analisis tentang pengaruh pendidikan sumber daya manusia terhadap tingkat pengangguran di Provinsi Jawa Tengah dan menyimpulkan bahwa kepadatan penduduk dan tingkat pendidikan mempengaruhi tingkat pengangguran. Dengan model Bayesian Hierarki, Ayuningtyas (2017) menggunakan variabel rata-rata lama sekolah dan pertumbuhan ekonomi sebagai variabel prediktor dalam penelitiannya. Soekarnoto (2014) dalam penelitiannya menyimpulkan bahwa pertumbuhan ekonomi melalui peningkatan Produk Domestik Regional Bruto (PDRB) memberikan pengaruh penting terhadap penurunan pengangguran terbuka di Provinsi Jawa Timur. Selain itu, Said (2017) dalam penelitiannya juga menyimpulkan bahwa pertumbuhan ekonomi memiliki pengaruh penting terhadap pengangguran di Kota Makassar.

Berdasarkan hasil beberapa penelitian di atas, variabel prediktor yang diduga mempunyai korelasi dengan tingkat pengangguran tersebut akan digunakan dalam penelitian ini. Variabel yang digunakan antara lain adalah kepadatan penduduk, persentase penduduk bekerja di sektor pertanian (sebagai pendekatan faktor usaha informal), rata-rata lama sekolah, dan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstant.