

**ESTIMASI MODEL REGRESI LINIER  
DENGAN PENDEKATAN BAYES  
(Studi Kasus Pada Data Curah Hujan di Seattle dan Portland)**

**SKRIPSI**

oleh:  
**DIANA RAHMAWATI**  
**NIM. 07610049**



**JURUSAN MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2011**

**ESTIMASI MODEL REGRESI LINIER  
DENGAN PENDEKATAN BAYES  
(Studi Kasus Pada Data Curah Hujan di Seattle dan Portland)**

**SKRIPSI**

oleh:  
**DIANA RAHMAWATI  
NIM. 07610049**



**JURUSAN MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2011**

**ESTIMASI MODEL REGRESI LINIER  
DENGAN PENDEKATAN BAYES  
(Studi Kasus Pada Data Curah Hujan di Seattle dan Portland)**

**SKRIPSI**

oleh:  
**DIANA RAHMAWATI**  
**NIM. 07610049**

Telah Diperiksa dan Disetujui untuk Diuji  
Tanggal: 20 Agustus 2011

Pembimbing I,

Pembimbing II,

Sri Harini, M.Si  
NIP. 19731014 200112 2 002

Dr. H. Munirul Abidin, M.Ag  
NIP. 19720420 200212 1 003

Mengetahui,  
Ketua Jurusan Matematika

Abdussakir, M.Pd  
NIP. 19751006 200312 1 001

**ESTIMASI MODEL REGRESI LINIER  
DENGAN PENDEKATAN BAYES  
(Studi Kasus Pada Data Curah Hujan di Seattle dan Portland)**

**SKRIPSI**

oleh :  
**DIANA RAHMAWATI**  
**NIM. 07610049**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Sains (S.Si)  
Tanggal: 12 September 2011

<b>Susunan Dewan Penguji</b>	<b>Tanda Tangan</b>
<b>1. Penguji Utama</b> : <b><u>Drs. H. Turmudi, M.Si</u></b> <b>NIP. 19571005 198203 1 006</b>	( )
<b>2. Ketua Penguji</b> : <b><u>Usman Pagalay, M.Si</u></b> <b>NIP. 19650414 200312 1 001</b>	( )
<b>3. Sekretaris Penguji</b> : <b><u>Sri Harini, M.Si</u></b> <b>NIP. 19731014 200112 2 002</b>	( )
<b>4. Anggota Penguji</b> : <b><u>Dr. H. Munirul Abidin, M.Ag</u></b> <b>NIP. 19800527 200801 1 012</b>	( )

**Mengesahkan,  
Ketua Jurusan Matematika**

**Abdussakir, M.Pd**  
**NIP. 19751006 200312 1 001**

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Diana Rahmawati  
NIM : 07610049  
Jurusan : Matematika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Agustus 2011

Yang membuat pernyataan,

Diana Rahmawati  
NIM. 07610049

## MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

إِنَّ اللَّهَ لَا يُغَيِّرُ مَا بِقَوْمٍ حَتَّى يُغَيِّرُوا مَا بِأَنْفُسِهِمْ

*Sesungguhnya Allah tidak merubah keadaan suatu kaum  
sehingga mereka*

*Merubah keadaan yang ada pada diri mereka sendiri*

*(Q.S. Ar, Ra'd: 11)*

*“We Can Take From Our Life up to What We Put To”*

## PERSEMBAHAN

**Ayah dan Ibu tercinta**, terimakasih atas setiap tetesan air mata dalam do'a mu untuk Ananda. Semoga Allah membalas semua kebaikan yang telah Ayah dan Ibu berikan pada Ananda selama ini karena hanya Allah yang bisa membalas kebaikan Ayah dan Ibu.

**Adik Atik Dwi Purwandari** yang telah perhatian, memberikan semangat, sehingga Penulis bisa menyelesaikan skripsi ini.

**Seseorang** yang telah menemani, memberikan semangat, perhatian dan bimbingannya, terimakasih atas semua yang engkau lakukan. Semoga hubungan ini adalah yang terbaik.

## KATA PENGANTAR



*Alhamdulillahirobbil'alamin*, teriring ucapan puja dan puji syukur kehadiran Allah SWT, yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “Estimasi Model Regresi Linier dengan Pendekatan Bayes (Studi Kasus pada Data Curah Hujan di Seattle dan Portland).

Shalawat serta salam semoga senantiasa dilimpahkan pada junjungan kita Nabi Muhammad SAW, keluarga serta sahabat-sahabatnya, yang telah membimbing kita dari zaman yang tidak beragama menuju zaman yang beragama yakni *Agama Islam*.

Penulisan skripsi ini dapat terselesaikan berkat bimbingan dan motivasi dari dosen pembimbing, bapak, ibu dosen serta bantuan dari semua pihak. Penulis menyadari bahwa dalam penulisan skripsi ini tidak terlepas dari banyak pihak. Oleh karena itu, tidak lupa penulis ucapkan banyak-banyak terima kasih kepada:

1. Prof. Dr. H. Imam Suprayogo selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Drs. Sutiman Bambang Sumitro, SU, D. Sc selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Abdussakir, M.Pd selaku Ketua Jurusan Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

4. Sri Harini, M.Si, selaku pembimbing I dan Dr. H. Munirul Abidin, M.Ag selaku pembimbing II. Terimakasih atas bimbingannya selama ini.
5. Segenap sivitas akademika Jurusan Matematika, terutama seluruh dosen, terima kasih karena telah memberikan ilmu pengetahuan kepada penulis selama dibangku perkuliahan.
6. Ayah dan Ibu tercinta atas doa, motivasi, kasih sayang serta segala pengorbanannya baik moril maupun spiritual dalam mendidik serta mendampingi penulis hingga dapat menyelesaikan tugas akhir ini.
7. Adik Atik Dwi Purwandari, terima kasih atas perhatian dan suportnya.
8. Prof. Dr. K.H. Ahmad Mudhor, S.H, Pengasuh Pesantren Luhur Malang yang dengan kesabarannya dan keikhlasannya membimbing penulis dalam mengarungi samudra ilmu
9. Teman-teman senasib seperjuangan mahasiswa matematika angkatan 2007 yang telah memberikan bantuan, motivasi, dan rasa kebersamaan yang terindah yang telah terukir selama masa perkuliahan.
10. Semua pihak yang telah berjasa dalam penulisan skripsi ini

Semoga Allah SWT membalas semua amal kebaikan yang telah mereka berikan kepada kami dan semoga skripsi ini dapat bermanfaat dan menambah khazanah keilmuan, Amin.

Malang, 20 Agustus 2011

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGAJUAN .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iv</b>
<b>HALAMAN PERNYATAAN KEASLIAN TULISAN .....</b>	<b>iv</b>
<b>MOTTO .....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>vii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>DAFTAR TABEL.....</b>	<b>xiii</b>
<b>ABSTRAK .....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian .....	4
1.4 Batasan Masalah .....	4
1.5 Kontribusi Penelitian .....	5
1.6 Metode Penelitian .....	5
1.7 Sistematika Penulisan.....	6
<b>BAB II KAJIAN PUSTAKA</b>	
2.1 <i>Outlier</i> .....	8
2.2 Analisis Regresi .....	9
2.3 Estimasi Parameter.....	11
2.4 Distribusi Normal .....	12
2.5 Model Regresi dalam Pendekatan Matriks .....	13
2.6 Pendekatan Bayes .....	14

2.7 Distribusi Prior.....	16
2.8 Distribusi Posterior.....	17
2.9 Pendekatan Klasik pada Inferensi Bayes.....	18
2.10 Bayesian <i>Marcov Chain Monte Carlo</i> (MCMC).....	20
2.11 WinBUGS.....	24
2.12 Kajian <i>Outlier</i> dan estimasi dalam Al-Qur'an.....	28
2.12.1 <i>Outlier</i> dalam Kajian Al-Qur'an.....	28
2.12.2 Estimasi dalam Kajian Al-Qur'an.....	30
<b>BAB III PEMBAHASAN</b>	
3.1 Membentuk Fungsi dari Model Regresi Linier.....	36
3.2 Membentuk Fungsi Likelihood.....	36
3.3 Menentukan Distribusi <i>Prior Noninformatif</i> .....	37
3.4 Menentukan <i>Joint Posterior</i> .....	39
3.5 Menentukan <i>Marginal Posterior</i> dari $\theta$ .....	40
3.6 Menentukan <i>Marginal Posterior</i> dari $\sigma$ .....	42
3.7 Penerapan pada Studi Kasus Data Curah Hujan di Seattle dan Portland yang Disimulasikan dengan Menggunakan Bantuan Paket Program WinBUGS Versi 1.4.....	44
3.7.1 Melakukan Pengecekan Data Apakah Terdapat <i>Outlier</i> atau Tidak.....	45
3.7.2 Menduga Parameter dengan Menggunakan <i>Ordinary Least Square</i> (OLS) dan Bayesian MCMC.....	49
3.7.2.1 Pendugaan Parameter dengan Ordinary Least Square.....	49
3.7.2.2 Pendugaan Parameter dengan Byaes MCMC.....	49
3.7.3 Analisis Perbandingan Model.....	51
3.8 Korelasi Al-Qur'an dengan Matematika.....	52

**BAB V PENUTUP**

4.1 Kesimpulan .....	58
4.2 Saran .....	59

**DAFTAR PUSTAKA**

**LAMPIRAN**



## DAFTAR GAMBAR

Gambar 2.1 Bivariate yang menunjukkan tiga titik data <i>outlier</i> .....	8
Gambar 2.2 Grafik distribusi normal .....	13
Gambar 2.3 <i>Graphical</i> eksponensial model .....	25
Gambar 2.4 Program WinBUGS dari Graphical model .....	25
Gambar 2.5 Program WinBUGS dari Graphical model lengkap dengan data dan inisialisasi parameter .....	26
Gambar 2.6 <i>Dynamic Trace Plot</i> dari eksponensial parameter .....	27
Gambar 2.7 Penduga parameter sebaran eksponensial .....	27
Gambar 2.8 <i>Kernel density</i> dari eksponensial parameter .....	27
Gambar 3.2 <i>Dynamic Trace</i> .....	50
Gambar 3.3 <i>Kernel Density</i> .....	50
Gambar 3.4 Menunjukkan tiga titik penyimpangan .....	54

## DAFTAR TABEL

Tabel 3.1	Data Curah Hujan .....	44
Tabel 3.2	Data yang terdapat <i>outlier</i> .....	45
Tabel 3.3	Analisis data regresi yang mengandung <i>outlier</i> .....	48
Tabel 3.4	Hasil Pendugaan Parameter dengan menggunakan MINITAB .....	49
Tabel 3.5	Hasil Pendugaan Parameter dengan menggunakan WinBUGS.....	51



## ABSTRAK

Rahmawati, Diana. 2011. **Estimasi Model Regresi Linier dengan Pendekatan Bayes (Studi Kasus pada Data Curah Hujan di Seattle dan Portland).**

Skripsi. Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Sri Harini, M.Si

(II) Dr. H. Munirul Abidin, M.Ag

Secara umum *outlier* dapat diartikan sebagai data yang tidak mengikuti pola umum model dan berjarak tiga kali simpangan baku atau lebih dari rata-rata (yaitu nol). *Outlier* merupakan salah satu faktor yang dapat mempengaruhi pendugaan parameter pada model regresi linier. Pada penelitian ini bertujuan untuk menduga parameter model regresi linier dengan pendekatan Bayes dan diharapkan dapat membantu para peneliti di dalam memilih metode penduga parameter untuk menghasilkan model terbaik.

Metode yang digunakan dalam menduga parameter dalam model regresi linier ini adalah metode Bayes yang akan diimplementasikan secara numerik melalui pendekatan *Markov Chain Monte Carlo* (MCMC) pada program WinBUGS. Metode Bayes memberikan hasil pendugaan yang lebih baik daripada pendugaan metode klasik. Hal ini disebabkan karena dalam metode klasik hanya berdasarkan informasi dari data sampel dan tidak mempertimbangkan informasi dari sebaran sebelumnya (*prior*).

Hasil dari penelitian ini didapatkan joint posterior  $\left(\frac{1}{\sigma^{(n+1)}}\right) \exp\left[-\frac{1}{2\sigma^2}\left(vs^2 + (\theta - \hat{\theta})' X' X (\theta - \hat{\theta})\right)\right]$ . Hasil joint posterior tersebut akan digunakan sebagai awalan dalam membangun MCMC untuk model regresinya. MCMC khususnya *Gibbs Sampler* yang digunakan disini akan menirukan proses Markov yang mencatat proses sekarang akan dipengaruhi satu step proses sebelumnya.

Keberhasilan peningkatan akurasi suatu model akan ditunjukkan pada suatu contoh kasus data dengan membandingkan hasil pemodelan dengan cara *Ordinary Least Square* (OLS) yang diimplementasikan melalui MINITAB dan dengan cara Bayes melalui WinBUGS. Dari hasil pemodelan tersebut menunjukkan metode Bayes MCMC lebih baik dibandingkan *Ordinary Least Square* (OLS). Hal ini disebabkan karena *Mean Square Error* (MSE) dalam OLS jauh lebih besar yaitu 17.29 daripada *Mean Square Error* (MSE) dalam Bayes MCMC yaitu 8.316.

**Kata Kunci:** Estimasi, *Outlier*, Bayesian *Markov Chain Monte Carlo*, Probabilitas Posterior

## ABSTRACT

Rahmawati, Diana. 2011. **The Estimation of Regression Linier Model Using Bayes Approach (A Case Study of Rainfall Data at Seattle and Portland)**. Thesis. Mathematic Department, Faculty of Science and Technology, Islamic State University Maulana Malik Ibrahim Malang.  
 Advisors: (I) Sri Harini, M.Si  
 (II) Dr. H. Muirul Abidin, M.Ag

Outlier, in common, is a statistical value that is outside other values and that the remainder is three times far from the standard deviation or more than the average (zero). Outlier is a factor that can affect the estimation on the parameter of regression linier model. This research aims to estimate the parameter of regression linier model using bayes approach and to help researchers in choosing parameter estimation methods to find an appropriate model.

The methode that is used to estimate the parameter of regression linier model is Bayes method that is implemented numerically through *Markov Chain Monte Carlo* (MCMC) approach at WinBUGS program. Bayes method gives a better estimation than klasik method does. It is because classic method gives information only from the samples and it does not consider the prior information (*prior*).

The result of this study is joint poterior  $\left(\frac{1}{\sigma^{(n+1)}}\right) \exp\left[-\frac{1}{2\sigma^2}\left(v s^2 + (\theta - \hat{\theta}) \mathbf{X} \mathbf{X} (\theta - \hat{\theta})\right)\right]$ . This result will be used as a base to bulid MCMC for regression model. MCMC especially *Gibbs Simpler* will follow the Markov process that is also influenced by its previous step.

The accuracy improvement of a model is accomplished can be seen from a case study that compares the modelling result of *Ordinary Least Square* (OLS) implemented by MINITAB and Bayes WinBUGS. This result shows that MCMC is better a better approach than *Ordinary Least Square* (OLS). It is because the *Mean Square Error* (MSE) of OLS is bigger than *Mean Square Error* (MSE) of Bayes MCMC, 17.29 compare with 8.316.

**Key Words:** Estimation, Outlier, Bayesian *Markov Chain Monte Carlo*, Probability Posterior.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Matematika merupakan ilmu yang mendasari berbagai macam ilmu yang lain misalkan ekonomi, kesehatan, pertahanan dan keamanan, budaya, sosial, politik dan agama. Sedangkan cabang ilmu matematika yang seringkali digunakan adalah statistik. Statistik yaitu metode atau ilmu yang mempelajari cara pengumpulan, pengolahan, penganalisisan, penafsiran, dan penarikan kesimpulan (Hasan, 2002:2).

Model regresi merupakan model yang cocok digunakan dalam menganalisis data penelitian yang melibatkan variabel *respon* (peubah terikat) dan variabel *explanatory* (peubah bebas). Model regresi ini mempunyai dua bentuk yaitu berbentuk linier dan tidak linier dalam parameteranya. Model yang linier dalam parameteranya adalah yang dapat didekati dengan teknik-teknik regresi berganda, seperti model-model polinom. Model yang tak linier dalam parameteranya dikatakan linier instrinsik bila suatu informasi dapat membuatnya linier. Kurva-kurva logaritma dan eksponensial termasuk golongan ini. Model yang tidak dapat dilinierkan melalui transformasi dikatakan nonlinier instrinsik dan analisis yang berhubungan disebut regresi tidak linier (Steel dan Torrie, 540:1993).

Suatu model regresi linier ataupun nonlinier tidak akan terlepas dari permasalahan sisaan. Sisaan (*residual*), dilambangkan  $\varepsilon_i$  dengan definisi

sebagai selisih antara nilai pengamatan ( $Y_i$ ) dan nilai ramalannya ( $\hat{Y}_i$ ), dengan  $i = 1, 2, \dots, n$  yang diperoleh dari persamaan regresi (Draper dan Smith, 1992:135). Sisaan tersebut sering disebut sebagai *outlier*. Secara umum *outlier* diartikan sebagai data yang tidak mengikuti pola umum suatu model dan secara kasar dapat diambil patokan yaitu sisanya yang berjarak tiga kali simpangan baku atau lebih rata-ratanya (yaitu nol).

Dalam Al Quran telah disinggung terkait dengan permasalahan *outlier*. Hal ini terdapat dalam Surat Al-Jin ayat 14:

وَأَنَا مِنَّا الْمُسْلِمُونَ وَمِنَّا الْقَاسِطُونَ فَمَنْ أَسْلَمَ فَأُولَئِكَ تَحَرَّوْا رَشَدًا ﴿١٤﴾

Artinya : “Dan sesungguhnya diantara kami ada orang-orang yang taat dan ada (pula) orang-orang yang menyimpang dari kebenaran. Barang siapa yang taat, maka mereka itu benar-benar telah memilih jalan yang lurus”.(QS. Al-Jin, 72:14).

Pada QS. Al-Jin ayat 14 tersebut dijelaskan bahwa terdapat suatu kaum jin yang taat dan patuh kepada Allah SWT dan ada pula para penyimpang. Dari penjelasan ayat diatas terdapat kata menyimpang, dalam ilmu statistik para penyimpang tersebut dianggap sebagai *outlier*. Karena *outlier* dapat diartikan sebagai data yang tidak mengikuti pola umum model atau data yang menyimpang (Sembiring,1995:62).

*Outlier* dapat terjadi karena kesalahan manusia, kesalahan instrument, perilaku curang, perubahan perilaku sistem atau kesalahan sistem, dan penyimpangan alami di dalam populasi. Kehadiran *outlier* sering kali berdampak buruk terhadap analisis data, karena *outlier* mampu

menyimpangkan tes-tes statistik yang didasarkan pada dua penaksir klasik yaitu rerata sampel dan kovariansi sampel.

Berawal dari kebutuhan analisis data untuk memprediksi suatu nilai bila diberikan nilai-nilai variabel  $x$  pada beberapa kasus maka metode regresi pun semakin berkembang. Adapun metode yang digunakan adalah metode klasik seperti *Ordinary Least Square* (OLS), namun metode ini dianggap kurang tepat untuk menganalisis sejumlah data yang tidak simetris. Karena pada metode klasik hanya melihat dua kelompok yang dibagi pada nilai tengahnya. Padahal ada kemungkinan kemiringan data terletak pada penduga kuartil tertentu.

Selanjutnya berkembanglah suatu metode Bayes. Keunggulan utama dalam penggunaan metode Bayes adalah penyederhanaan dari cara klasik yang penuh dengan integral untuk memperoleh model marginal. Disamping itu, metode Bayes memberikan hasil pendugaan yang lebih baik daripada pendugaan dalam metode klasik. Karena di dalam metode klasik dalam pendugaan parameternya hanya berdasarkan informasi dari data sampel, dimana ukuran sampel sangat berpengaruh terhadap hasil pendugaan.

Dalam metode Bayes selain menggunakan informasi dari data sampel juga dipertimbangkan informasi dari sebaran prior untuk mendapatkan sebaran posterior, sehingga hasil pendugaan dalam metode Bayes akan jauh lebih baik. Pada metode Bayes digunakan pendekatan algoritma komputasional *Markov Chain Monte Carlo* (MCMC), khususnya teknik *Gibbs Sampler* yang diimplementasikan pada paket program **WinBUGS** 1.4.

Dari latar belakang diatas, dalam skripsi ini peneliti akan mengkaji *“Estimasi Model Regresi Linier dengan Pendekatan Bayes (Studi Kasus pada Data Curah Hujan di Seattle dan Portland).*

## 1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah bagaimana menduga parameter model regresi linier dengan metode Bayes pada kasus data curah hujan di Seattle dan Portland ?

## 1.3 Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah untuk menduga parameter model regresi linier dengan metode Bayes pada kasus data curah hujan di Seattle dan Portland.

## 1.4 Batasan Masalah

Dalam penelitian ini, pembahasan masalah akan dibatasi mengenai:

1. Distribusi sampel yang digunakan adalah distribusi normal univariat
2. Distribusi prior yang digunakan adalah distribusi prior noninformtaif
3. Sisaan random berdistribusi normal  $N(0, \sigma^2)$
4. Pada pendekatan Bayes MCMC digunakan bantuan WinBUGS 1.4

## 1.5 Kontribusi Penelitian

### Bagi Instansi

- a. Peningkatan kualitas keilmuan fakultas dengan adanya penelitian dan pengembangan penelitian
- b. Untuk menambah kepustakaan pengetahuan keilmuan dalam bidang ilmu matematika khususnya pada bidang ilmu regresi

### Bagi Pembaca

- a. Dapat membantu para peneliti di dalam memilih metode penduga parameter untuk menghasilkan model terbaik.
- b. Sebagai referensi apabila ingin mengembangkan ilmu regresi

## 1.6 Metode Penelitian

Penulisan skripsi ini dilakukan dengan pendekatan studi literatur. Studi literatur dilakukan untuk mengkonstruksi model dan mengestimasi model regresi linier dengan pendekatan Bayes.

Beberapa langkah yang harus dilakukan untuk mengestimasi model regresi linier dengan pendekatan Bayes, adalah sebagai berikut:

1. Menentukan Fungsi dari Model Regresi Sederhana
2. Membentuk fungsi likelihood  $L(\theta, \sigma | y)$  dari fungsi kepadatan probabilitas

$$f(x_i | \theta, \sigma)$$

3. Menentukan distribusi *prior noninformatif*
4. Membentuk Joint Posterior
5. Menentukan Marginal Posterior  $\theta$

6. Menentukan Marginal Posterior  $\sigma$
7. Penerapan pada Studi Kasus pada Data Curah Hujan di Seattle dan Portland yang disimulasikan dengan menggunakan bantuan paket program *Windows Bayesian Inference Using Gibbs Sampling (WinBUGS)* versi 1.4.

### 1.7 Sistematika Penulisan

Dalam penulisan tugas terakhir ini, penulis menggunakan sistematika penulisan yang terdiri dari empat bab, dan masing-masing bab dibagi dalam subbab dengan sistematika penulisan sebagai berikut:

**BAB I** : Pendahuluan, berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kontribusi penelitian, metode penelitian, dan sistematika penelitian

**BAB II** : Kajian Pustaka, berisi hal-hal yang mendasar dalam teori yang dikaji, meliputi: *Outlier*, analisis regresi, estimasi parameter, distribusi normal, model regresi dalam pendekatan matriks, pendekatan Bayes, distribusi *prior*, distribusi *posterior*, pendekatan klasik pada inferensi Bayes, Bayesian *Markov Chain Monte Carlo* (MCMC), inferensi Bayes, kajian *outlier* dan estimasi dalam Al-Quran.

**BAB III** : Pembahasan, Pembahasan pada bab ini berisi uraian tentang cara mengestimasi model regresi linier dengan pendekatan Bayes yang kemudian akan diterapkan pada studi kasus data curah hujan di Seattle dan Portland yang akan disimulasikan

dengan menggunakan bantuan paket program WinBUGS versi 1.4 dan korelasi Al-Quran dengan Matematika.

**BAB IV** : Penutup, berisi kesimpulan akhir penelitian dan saran untuk pengembangan penelitian selanjutnya yang lebih baik.

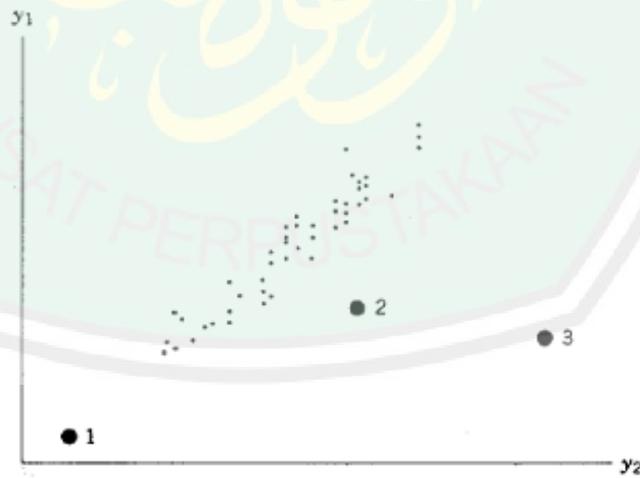


## BAB II

### KAJIAN PUSTAKA

#### 2.1 *Outlier*

Belum ada patokan yang disepakati oleh para statistikawan kapan suatu pengamatan dapat dikategorikan sebagai *outlier*. Secara umum, *outlier* ialah data yang tidak mengikuti pola umum model dan secara kasar, dapat diambil patokan yaitu yang sisanya berjarak 3 simpangan baku atau lebih dari rata-ratanya (yaitu nol) (Sembiring, 1995:62). Menurut Draper (1992:146), sisaan yang merupakan *outlier* adalah yang nilai mutlaknya jauh lebih besar dari pada sisaan-sisaan lainnya dan bisa jadi terletak tiga atau empat kali simpangan baku atau lebih jauh dari rata-rata sisanya.



Gambar 2.1 Bivariat yang menunjukkan tiga titik data *outlier*

(Alvin C, 2002: 103).

Pada gambar 2.1 di atas dapat dilihat bahwa terdapat tiga titik data yang terpisah sangat jauh dari data lainnya yaitu data 1, 2, dan 3 dan bisa jadi ketiga titik data tersebut terletak tiga atau empat kali simpangan bakunya. Ketiga titik itulah yang disebut sebagai *outlier*.

## 2.2 Analisis Regresi

Istilah regresi diperkenalkan pertama kali oleh Francis Galton, dalam makalahnya yang berjudul *Family Likeness in Stature*. Analisis regresi adalah teknik analisis yang mencoba menjelaskan bentuk hubungan antara peubah-peubah yang mendukung sebab akibat. Prosedur analisisnya didasarkan atas distribusi probabilitas bersama peubah-peubahnya. Bila hubungan ini dapat dinyatakan dalam persamaan matematika, maka dapat dimanfaatkan untuk keperluan-keperluan yang lain, misalnya peramalan. Secara umum, dapat dikatakan bahwa analisis regresi berkenaan dengan studi ketergantungan suatu variabel, yaitu variabel tak bebas (*dependent variable*) dan variabel bebas (*independent variable*), (Firdaus, 2004:22).

Menurut Supranto (1994:262), hubungan fungsi antara variabel X (variabel bebas) dan Y (variabel terikat) tidak selalu bersifat linier, akan tetapi bisa juga nonlinier. Diagram pencar dari hubungan yang linier akan menunjukkan suatu pola yang dapat didekati dengan garis lurus, sedangkan yang bukan linier harus didekati dengan garis lengkung.

Analisis regresi linier dapat dibedakan menjadi dua, yaitu:

1. Analisis regresi sederhana (*simple regression analysis*) atau regresi dua variabel, yang mempelajari ketergantungan satu variabel tak bebas hanya pada satu variabel bebas. Adapun model regresi sederhananya adalah:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ dimana } i = 1, 2, 3, \dots, n \quad (2.1)$$

Keterangan:

$y_i$  = variabel terikat (*dependent variable*)

$x_i$  = variabel bebas (*independent variable*)

$\beta_0$  = parameter konstanta/ intersept regresi yang tidak diketahui nilainya dan akan diestimasi

$\beta_1$  = parameter konstanta/ intersept regresi yang tidak diketahui nilainya dan akan diestimasi

$\varepsilon$  = variabel galat/kesalahan regresi, dengan  $\varepsilon \sim N(0; \sigma^2)$

2. Analisis regresi berganda (*multiple regression analysis*) atau regresi lebih dari dua variabel, yang mempelajari ketergantungan suatu variabel terikat pada lebih dari satu variabel bebas. Adapun model regresi bergandanya adalah:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i, \text{ dimana } i = 1, 2, 3, \dots, n \quad (2.2)$$

Keterangan:

$y_i$  = variabel terikat (*dependent variable*)

$x_i$  = variabel bebas (*independent variable*)

$\beta_0$  = parameter konstanta/ intersept regresi yang tidak diketahui nilainya dan akan diestimasi

$\beta_1$  = parameter konstanta/ intersept regresi yang tidak diketahui nilainya dan akan diestimasi

$\varepsilon$  = variabel galat/kesalahan regresi, dengan  $\varepsilon \sim N(0; \sigma^2)$

$k$  = banyaknya variabel bebas

(Firdaus, 2004:25)

### 2.3 Estimasi Parameter

Menurut Hasan (2001:111), pendugaan (estimasi) merupakan proses yang menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi yang tidak diketahui. Pendugaan merupakan suatu pernyataan mengenai parameter populasi yang diketahui berdasarkan populasi dari sampel dalam hal ini sampel random yang diambil dari populasi yang bersangkutan. Jadi dengan pendugaan ini, keadaan parameter populasi dapat diketahui. Menurut Yitnosumarto (1990 : 211-212), penduga (estimator) adalah anggota peubah acak dari statistik yang mungkin untuk sebuah parameter (anggota peubah diturunkan). Besaran sebagai hasil penerapan penduga terhadap data dari semua contoh disebut nilai duga (estimase).

Statistik merupakan sekumpulan konsep dan metode untuk mengumpulkan data, menyajikan data dalam bentuk yang mudah difahami, menganalisis data dan mengambil suatu kesimpulan berdasarkan hasil analisis data dalam situasi yang memiliki ketidakpastian dan variasi. Karena statistika

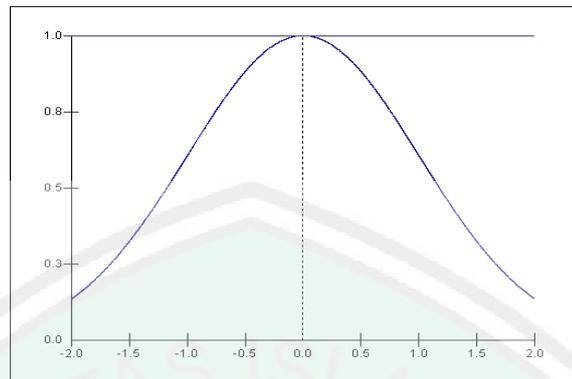
bersandar pada cara berfikir probabilistik, maka hasil pengolahan data yang menggunakan metode statistika bukanlah hasil pasti, tetapi merupakan hasil taksiran adanya ketidakpastian dari variansi yang terjadi dalam fenomena tertentu. Teknik pengambilan tentang suatu parameter meliputi pendugaan (estimasi) parameter dan pengujian hipotesis.

Salah satu aspek penting dalam statistik inferensia adalah pendugaan *parameter populasi*. Misalnya,  $\mu$  dan  $\sigma^2$  yang diduga dari statistik sampel  $\bar{x}$  dan  $s^2$ . Dengan demikian kesimpulan yang didapatkan merupakan kesimpulan tentang populasi yang dipelajari berdasarkan contoh atau sebagian dari populasi tersebut.  $\bar{x}$  dan  $s^2$  merupakan suatu peubah acak yang besarnya beragam dari satu contoh ke contoh lain serta memiliki sebaran statistik yang sesuai dengan sebaran induknya (Harini, 2008: 225). Adapun teknik pendugaan (estimasi) digolongkan menjadi dua yaitu estimasi titik dan estimasi interval.

#### 2.4 Distribusi Normal

Distribusi yang penting dalam statistika ialah distribusi normal atau sering disebut distribusi Gauss.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (2.4)$$



Gambar 2.2 Grafik distribusi normal

Distribusi ini mempunyai rata-ran  $\mu$  dan variansi  $\sigma^2$ . Grafiknya berbentuk gema yang simetris. Suatu peubah acak  $Y$  yang berdistribusi normal dengan rata-ran  $\mu$  dan simpangan baku  $\sigma$  sering disingkat dengan lambang  $Y \sim N(\mu, \sigma^2)$ .

## 2.5 Model Regresi dalam Pendekatan Matriks

Model yang paling sederhana adalah model regresi linier. Model regresi linier sederhana terdiri dari satu variabel. Model tersebut dapat digeneralisasikan menjadi lebih dari satu atau dalam  $k$  variabel. Persamaan bagi model regresi linier dengan  $k$  variabel diberikan sebagai berikut:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.5)$$

Dengan sejumlah  $n$  data observasi maka model linier ini dapat ditulis dalam bentuk matrik sebagai berikut

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.6)$$

Sehingga model ini dapat disederhanakan sebagai berikut

$$Y = X\beta + \varepsilon \quad (2.7)$$

Dengan:

$Y$  = vektor  $n \times 1$  dari variabel terikat

$X$  = matriks peubah bebas  $n \times p$

$\beta$  = parameter koefisien regresi

$\varepsilon$  = variabel galat regresi

(Aziz, 2007:21-22).

## 2.6 Pendekatan Bayes

Bayes memperkenalkan suatu metode dimana kita perlu mengetahui bentuk distribusi awal (*prior*) dari populasi yang dikenal dengan metode bayes. Sebelum menarik sampel dari suatu populasi terkadang kita peroleh informasi mengenai parameter yang akan diestimasi. Informasi ini kemudian digabungkan dengan informasi dari sampel untuk digunakan dalam mengestimasi parameter populasi. Dalam Metode Estimasi Bayesian, yang perlu diperhatikan yaitu parameter  $\mu$ . Parameter  $\mu$  mempunyai distribusi probabilitas  $P(\mu)$  yang merupakan tingkat kepercayaan awal tentang parameter  $\mu$  sebelumnya pengamatan dilakukan, yang dinamakan distribusi prior  $\mu$ .

Teorema umum bayes adalah (Box,1973):

$$P(\mu|y) = \frac{P(y|\mu) P(\mu)}{P(y)}$$

Dimana  $P(\mu|y)$  = distribusi posterior  $\mu$  dan  $P(y|\mu) P(\mu)$  pada umumnya tidak diketahui biasanya hanya distribusi prior dan fungsi likelihoodnya yang dinyatakan.

Rumus Bayes juga dapat ditulis (Box,1973):

$$P(\theta|y) \sim L(\theta, y) P(\theta)$$

Pada penilaian metode ini digunakan estimasi informasi prior (dulu) yang mana vektor parameter  $\theta = (\beta', \sigma)'$ . Dimana probabilitas fungsi kepadatan dari  $Y, f(y, \theta)$  diasumsikan sebagai normal *multivariat* dan menggabungkan pengetahuan awal tentang  $\theta$  yang dikaitkan dengan fungsi kepadatan  $f(\theta)$  dari  $\theta$ . Hal ini bertujuan untuk membuat kesimpulan yang didasarkan atas fungsi kepadatan untuk  $\theta$  pada  $Y = y$ . Dengan menggunakan teorema bayes dapat diperoleh fungsi kepadatan posterior dari  $\theta$  adalah

$$\begin{aligned} f(\theta|y) &= \frac{f(y|\theta)f(\theta)}{f(y)} \\ &= \frac{f(y, \theta)f(\theta)}{f(y)} \\ &= cf f(y, \theta)f(\theta) \end{aligned} \tag{2.8}$$

Dimana  $c$  tidak melibatkan  $\theta$ . Asumsi umum untuk  $\beta$  dan  $\sigma$  adalah distribusi prior independen.

## 2.7 Distribusi Prior

Permasalahan utama dalam pendekatan bayes ini adalah memilih distribusi prior  $g(\theta)$  yang menunjukkan ketidakpastian tentang parameter  $\theta$  yang tidak diketahui. Distribusi prior dapat dipilih melalui data masa lalu yang telah ada dan distribusi prior ini bisa disebut dengan distribusi prior “*data based*” (DB), jika data masa lalu tidak tersedia. Distribusi prior dipilih berdasarkan kepercayaan peneliti, dan distribusi prior jenis ini disebut “*non data based*” (NDB).

Adapun pengelompokan distribusi prior dilihat dari sudut pandang tertentu :

1. Berkaitan dengan bentuk distribusi hasil identifikasi pola datanya
  - a. Distribusi prior sekawan (*conjugate prior*) mengacu pada acuan analitis model terutama dalam pembentukan fungsi likelihoodnya. Sehingga dalam menentukan prior sekawan selalu dipikirkan mengenai penentuan pola distribusi prior yang mempunyai bentuk sekawan dengan fungsi densitas pembangun fungsi likelihoodnya.
  - b. Distribusi prior tidak sekawan (*non conjugate prior*) apabila pemberian prior pada suatu model tidak mengindahkan pola pembentuk likelihoodnya.
2. Berkaitan dengan penentuan masing-masing parameter pada distribusi prior tersebut dikelompokkan menjadi :
  - a. Distribusi *prior informatif*, yaitu distribusi yang mengacu pada pemberian parameter dari distribusi prior yang telah dipilih baik prior

yang dipilih sekawan atau tidak pemberian nilai parameter pada distribusi prior ini akan sangat mempengaruhi bentuk distribusi posterior yang akan di dapat pada informasi data yang akan diperoleh.

b. Distribusi *prior noninformative*, distribusi yang pilihannya tidak didasarkan pada data yang ada atau prior yang tidak mengandung informatif tentang  $\theta$ . *Prior Jeffrey* adalah pendekatan dari non informatif untuk 1 parameter.

3. Distribusi prior yang dibedakan atas ada dan tidaknya bentuk tetap untuk setiap variabel acak  $t$  yaitu prior proper dan *prior improper* atau *prior quasi*.

Prior ini timbul bila  $g(\theta)$  bukan distribusi probabilitas yaitu  $g(\theta) \geq 0$  tetapi  $\int g(\theta) d\theta \neq 1$ .

## 2.8 Distribusi Posterior

Selain distribusi Prior distribusi yang harus diketahui dalam pendekatan bayes adalah distribusi posterior. Distribusi ini berkaitan dengan penentuan masing-masing parameter pada pola distribusi prior tersebut. Distribusi prior informatif mengacu pada pemberian parameter dari distribusi prior yang telah dipilih. Baik prior yang dipilih sekawan maupun tidak, pemberian nilai parameter pada distribusi prior ini akan sangat mempengaruhi bentuk distribusi posterior yang akan di dapatkan pada informasi data yang diperoleh. Untuk mendapatkan distribusi posterior dari  $\beta$ , distribusi bersama dari  $p$  dan sampel yang akan diambil harus dihitung terlebih dahulu.

posterior  $\sim$  likelihood  $\times$  prior

Distribusi posterior untuk  $\theta$ , jika pengamatan  $y$  telah diambil merupakan gabungan dari informasi prior dan informasi data yang ditulis  $h(y|\theta)$  sehingga:

$$h(y|\theta) = \frac{P(y|\theta)}{\int P(y|\theta) d\theta} = \frac{P(\theta)P(y|\theta)}{\int P(\theta)P(y|\theta) d\theta}$$

Distribusi Posterior adalah distribusi prior yang disesuaikan dengan informasi sampel. Secara umum distribusi posterior dirumuskan sebagai berikut (Bain and Engelhardt :1992) :

$$f_{\theta|x}(\theta) = \frac{f(x_1, x_2, \dots, x_n|\theta)P(\theta)}{\int f(x_1, x_2, \dots, x_n|\theta)P(\theta)d\theta}$$

Distribusi  $f(x_1, x_2, \dots, x_n|\theta)P(\theta)$  merupakan fungsi likelihood dari  $\theta$  dan  $P(\theta)$  merupakan distribusi prior dari  $\theta$  sehingga dapat ditulis :

$$\text{Distribusi posterior} = \frac{(\text{likelihood})(\text{distribusi prior})}{\int (\text{likelihood})(\text{distribusi prior})}$$

(Hogg and Craig, 1970)

## 2.9 Pendekatan Klasik pada Inferensi Bayes

Metode penduga klasik mendasarkan semata-mata pada informasi yang dikandung dalam contoh. Metode tersebut pada dasarnya menafsirkan peluang sebagai frekuensi relative. Misalnya ketika memperoleh selang kepercayaan 95% bagi  $\mu$ , kita menafsirkan pernyataan peluang  $P(-1.96 < Z < 1.96) =$

0.95. dalam pengertian bahwa bila percobaan itu diulang berkali-kali maka 95% diantara  $Z$  yang diperoleh akan terletak diantara  $-1.96 < Z < 1.96$ . peluang sejenis ini, yang dapat ditafsirkan dalam pengertian frekuensi akan disebut sebagai peluang objektif. Pendekatan bayes dalam terhadap metode penduga statistik menggabungkan informasi yang dikandung dalam contoh dengan informasi lain yang telah tersedia sebelumnya.

Dalam pendekatan klasik, kita akan mengambil suatu contoh acak berukuran  $n$  dan kemudian mensubstitusikan informasi yang dikandung oleh contoh ke dalam suatu fungsi penduga/ fungsi keputusan. Jadi misalnya saja nilai dugaan bagi proporsi keberhasilan dalam suatu percobaan binom adalah  $\hat{p} = \frac{x}{n}$  dan nilai dugaan bagi parameter  $\mu$  dari suatu populasi normal adalah  $\bar{X}$ . Teknis Bayes menggunakan sebaran awal bersama-sama dengan bukti yang dikandung oleh contoh untuk menghitung sebaran posterior bagi parameter  $\theta$ . Penarikan kesimpulan mengenai parameter populasi selanjutnya di dasarkan pada sebaran posterior ini. Misalnya saja, nilai tengah sebaran posterior ini dapat digunakan sebagai nilai dugaan titik bagi  $\theta$  (Walpole, 1995:277).

Dari segi asumsi statistikawan klasik memandang bahwa parameter populasi mempunyai harga tertentu yang tidak diketahui sehingga pernyataan probabilitas tentang parameter populasi tidak mempunyai arti. Dalam pendekatan klasik estimasi parameter telah ditentukan, akan tetapi tidak diketahui. Sebelum data dikumpulkan untuk level  $(1 - r)$  selang kepercayaan (random) akan berisi parameter dengan probabilitas  $(1 - r)$ . Setelah data dikumpulkan penghitungan selang kepercayaan baik yang berisi estimasi

parameter maupun tidak, biasanya kita tidak pernah mengetahui yang mana yang benar. Dan sebaliknya, inferensi Bayesian parameter yang tidak diketahui  $\beta_\theta$  nya akan diterapkan sebagai variabel random dan variabel random ini dalam perhitungannya menurun. Selang kepercayaan deterministik dengan probabilitas  $(1 - r)$ .

Anggap kepadatan kondisional data vektor  $(x, y)$  diberikan  $\beta_\theta$  yang dinotasikan dengan  $\pi(x|\beta_\theta)$ , dan andaikan distribusi prior parameter  $\beta_\theta$  ditetapkan dengan populasi  $\pi$ . Kepadatan dari data vektor dan peremetnya akan menjadi  $\pi(data|\beta_\theta) \pi(\beta_\theta)$ , dan kepadatan posterior dari data yang diberikan adalah (dengan teorema Bayes)  $\pi(\beta_\theta|data) \propto \pi(data|\beta_\theta) \pi(\beta_\theta)$ . Sekarang misalkan  $A(X)$  adalah himpunan kepercayaan (sebuah subset dari ruang parameter bergantung pada data, tetapi parameternya tidak diketahui). Definisi yang tepat untuk level  $(1 - r)$  bayesian adalah

$$P[\beta_\theta \in A(X)|X = x] = 1 - r$$

Dimana  $\beta_\theta$  adalah acak, dan probabiliitas dari definisi diatas dapat dihitung dengan menggunakan kepadatan posterior  $\pi(\beta_\theta|data)$ .

### 2.10 Bayesian *Marcov Chain Monte Carlo* (MCMC)

Menurut Pereira (1999), metode Bayes memberikan hasil pendugaan yang lebih baik daripada pendugaan dalam metode klasik. Hal ini disebabkan pendugaan parameter hanya berdasarkan informasi dari data sampel, dimana ukuran sampel sangat berpengaruh terhadap hasil pendugaan. Dalam metode Bayes selain menggunakan informasi dari data sampel juga dipertimbangkan

informasi dari sebaran prior untuk mendapatkan sebaran posterior, sehingga hasil pendugaan dalam metode Bayes akan lebih baik. Namun pada kenyataannya, sebaran prior tidak mudah ditentukan dan sebaran posterior menjadi sulit diperoleh sehingga metode Bayes sulit diselesaikan secara analitik. Untuk mengatasi masalah tersebut, maka dikembangkan teknik simulasi sehingga metode Bayes mudah diselesaikan.

Teknik simulasi yang biasa digunakan dalam metode Bayes adalah metode *Markov Chain Monte Carlo* (MCMC). menurut Scollnik (1996), metode MCMC merupakan metode simulasi untuk mendapatkan data sampel suatu peubah acak dengan teknik sampling berdasarkan sifat rantai markov. Salah satu teknik dalam metode MCMC yang terkenal adalah *Gibbs Sampler*. Dalam melakukan proses simulasi, Gibbs Sampler menggunakan sebaran bersyarat untuk membangkitkan data sampel peubah acak.

Metode Bayes merupakan metode yang menggabungkan informasi prior dengan pengamatan di dalam percobaan sehingga menghasilkan sebaran posterior. Sebaran posterior kemudian digunakan untuk memperbaharui sebaran prior melalui data pengamatan (Pereire,1999).

Apabila  $\theta$  merupakan suatu nilai peubah acak dengan sebaran peluang  $f(\theta)$ , maka  $f(\theta)$  sering disebut sebagai sebaran sebaran awal atau sebaran prior. Selanjutnya sebaran prior  $f(\theta)$  digunakan bersama sebaran bersyarat  $f(x|\theta)$  dalam sebaran gabungan sampel  $f(x, \theta) = f(x|\theta)f(\theta)$ . Sebaran bersyarat pada dalam metode Bayes juga dapat didefinisikan sebagai penggabungan fungsi likelihood dan sebaran prior. Untuk menentukan nilai

duga parameter dari sebaran posterior  $\theta$  digunakan metode *Markov Chain Monte Carlo* (MCMC). metode MCMC ini cukup efektif untuk menentukan nilai duga parameter dari sebaran posterior yang sangat kompleks dan cukup sulit jika diselesaikan dengan metode lain (Pereira, 1999).

Jika  $X$  merupakan variabel acak  $(X_t)$  dimana  $t \in T$  merupakan indeks waktu atau deretan. Sebuah proses stokastik memperlihatkan sifat markov jika kejadian pada saat  $t+1$  yaitu peubah acak  $X_{t+1}$  hanya dipengaruhi oleh kejadian satu langkah sebelumnya. Secara matematis sifat rantai markov dapat ditulis sebagai berikut:

$$P\{X_{t+1} = j | X_{t+1} = i\} = P\{X_{t+1} = j | X_t = i, X_{t-1} = i_1, X_{t-2} = i_2, \dots, X_0 = i_0\} \quad (2.9)$$

Untuk  $t = 1, 2, \dots, n$  dan setiap deretan  $j, i, i_1, \dots, i_t$ .

$$\text{Peluang bersyaratnya: } P\{X_{t+1} = j | X_{t+1} = i\} = P_{ij} \quad (2.10)$$

Disebut peluang transisi satu langkah sedangkan peluang transisi  $n$  tahap

$$\text{adalah } P\{X_{t+1} = j | X_{t+1} = i\} = P_{ij}^{(n)} \quad (2.11)$$

Rantai markov dikatakan memiliki sebaran stasioner  $\pi(x)$  jika

$$P\{X_{t+1} = j | X_{t+1} = i\} = P\{X_1 = j | X_0 = i\} = P_{ij} \quad (2.12)$$

Simulasi Monte Carlo merupakan suatu pendekatan untuk menduga fungsi sebaran dari peubah acak  $\{X_t\}$ . Metode simulasinya adalah Gibbs Sampler yaitu metode yang menggunakan sebaran bersyarat penuh yang dihubungkan dengan sebaran stasioner  $\pi(x)$  (Scollnik, 1996).

Misalkan  $\pi(x_1, x_2, \dots, x_k)$  merupakan sebaran gabungan dari peubah  $X_j$  yang bersesuaian dengan sebaran stasioner  $\pi(x)$ , maka  $\pi(x_j)$  merupakan sebaran marginal dan  $\pi(x_j | x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k)$  merupakan sebaran bersyarat dari variabel  $X_j$ , dimana  $j=1, 2, \dots, k$ . Sehingga proses Gibbs Sampler dilakukan dengan cara membangkitkan sampel dari sebaran bersyarat penuh dengan mengikuti langkah-langkah sebagai berikut:

1. Memilih nilai awal

Nilai awal yang diberikan ini adalah sebarang nilai yang sesuai dengan peubah acak  $X_1, X_2, \dots, X_k$  dan akan digunakan untuk menduga  $x^{(1)}$ , dan  $x^{(1)}$  akan digunakan untuk menduga  $x^{(2)}$ , dan seterusnya sampai  $x^{(n)}$ .

2. Dilakukan simulasi pengambilan sampel dari peubah acak berdasarkan sebaran bersyarat penuh

$$X_1^{(0)} \sim \pi(x_1 | x_2^{(i-1)}, \dots, x_k^{(i-1)})$$

$$X_2^{(0)} \sim \pi(x_2 | x_1^{(i)}, x_3^{(i-1)}, \dots, x_k^{(i-1)})$$

$$X_3^{(0)} \sim \pi(x_3 | x_1^{(i)}, x_2^{(i)}, x_4^{(i-1)}, \dots, x_k^{(i-1)})$$

.

.

.

$$X_k^{(0)} \sim \pi(x_k | x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i-1)})$$

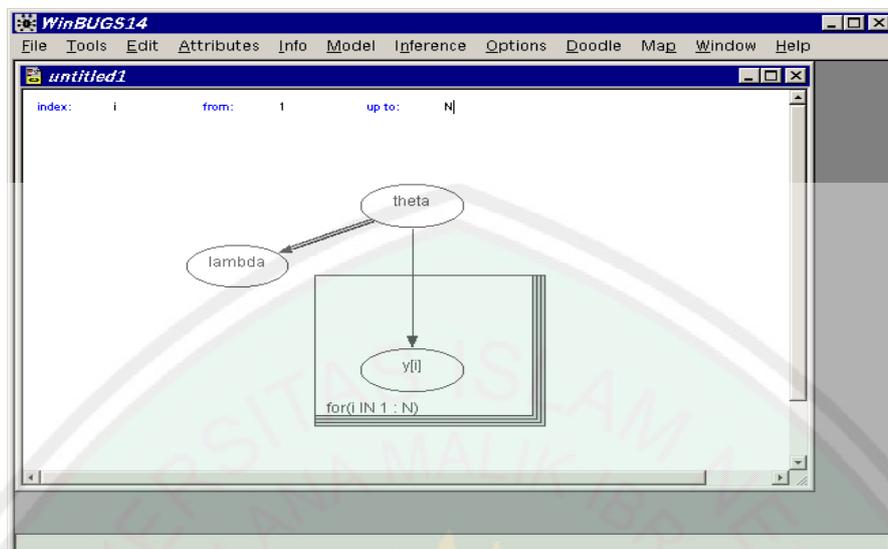
Sehingga  $X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \dots, X_k^{(i)})$

Simulasi ini dilakukan mulai  $i = 1$  hingga  $n$  iterasi yang diinginkan. Ketika  $n \sim \infty$ , maka nilai  $X^{(n)}$  dari masing-masing sebaran bersyarat penuh akan tampak sangat acak.

### 2.11 WinBUGS

WinBUGS adalah sebuah paket program yang dirancang khusus untuk memfasilitasi pemodelan data dengan basis Bayesian dengan implementasi *Markov Chain Monte Carlo* (MCMC). Nama WinBUGS diambil dari isi paket programnya yang dikembangkan berdasarkan pada metode *Gibbs sampler* dan dibuat untuk dapat di *running* di dalam sistem operasi komputer Windows. Jadi inti dan pengertian nama WinBUGS adalah *Bayesian Using Gibbs Sampler* (BUGS) dalam Windows.

Langkah pertama pemrograman WinBUGS yaitu dengan membentuk struktur pemodelan grafik dalam *doodle* WinBUGS dengan nama-nama node yang bersesuaian dengan nama variable dalam model. Node merepresentasikan variable dari model. Terdapat tiga type dari sebuah node yaitu *stochastic*, *logical* dan *constant*. Node dengan bentuk ellip akan berarti bertipe stokastik atau logikal, sedangkan apabila node tersebut berbentuk kotak berarti node tersebut bertipe constant. Gambar 2.1 contoh sebuah doodle dengan tipe stochastic dan logical untuk merepresentasikan estimasi model/distribusi data yang berdistribusi eksponensial dengan parameter tunggal.



Gambar 2.3 *Graphical* eksponensial model

Gambar 2.3 merupakan *graphical model* yang dilengkapi dengan plate. Plate atau bingkai ini memuat index  $i$  dalam node  $y[i]$ . Jika *graphical model* telah selesai dibuat, maka akan dibuatkan window baru yang memuat tampilan program code untuk estimasi densitas *Exponential* sebagai aplikasi WinBUGS yang dibuat secara *automatic programming* berdasarkan model grafik yang telah dibuat di atas.

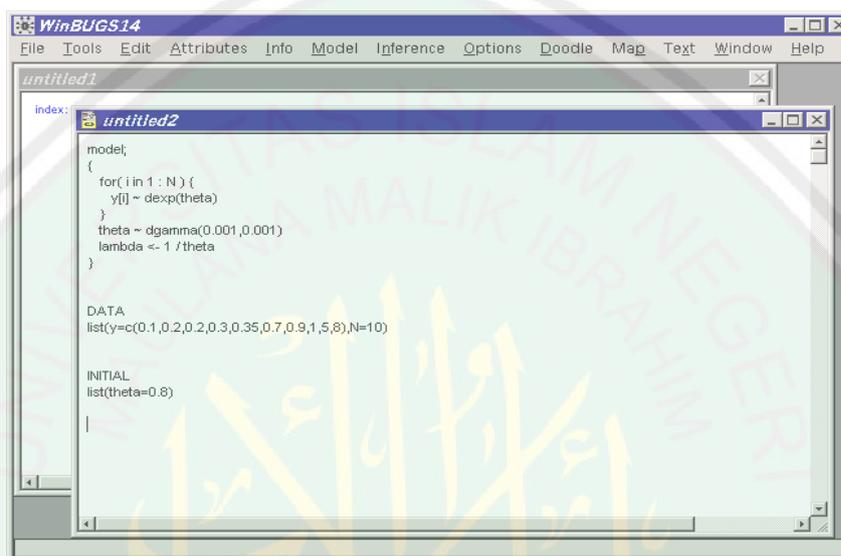
```

model;
{
  for( i in 1 : N ) {
    y[i] ~ dexp(theta)
  }
  theta ~ dgamma(0.001,0.001)
  lambda <- 1 / theta
}

```

Gambar 2.4 Program WinBUGS dari Graphical model

Untuk menjalankan program WinBUGS di atas, perlu memberikan data masukan dan nilai inialisasi proses iteratif MCMC-nya. Adapun data dan data inialisasinya diawali dengan *reserve word* 'list'.



```

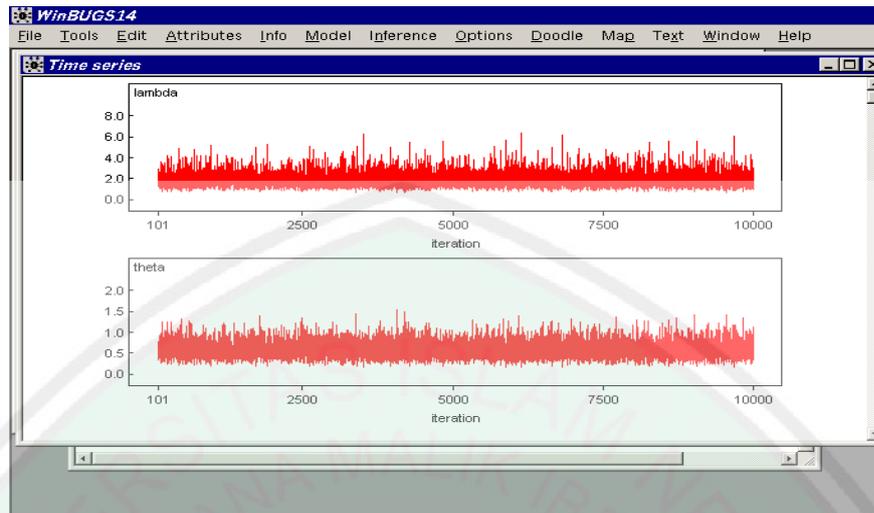
WinBUGS14
File Tools Edit Attributes Info Model Inference Options Doodle Map Text Window Help
untitled1
Index:
untitled2
model;
{
  for( i in 1 : N ) {
    y[i] ~ dexp(theta)
  }
  theta ~ dgamma(0.001,0.001)
  lambda <- 1 / theta
}

DATA
list(y=c(0.1,0.2,0.2,0.3,0.35,0.7,0.9,1.5,8),N=10)

INITIAL
list(theta=0.8)
  
```

Gambar 2.5 Program WinBUGS dari *Graphical model* lengkap dengan data dan inialisasi parameter

Dari program tersebut, simulasi dapat dijalankan sehingga diperoleh nilai konvergen bagi parameter yang diduga. Kekonvergenan dapat diketahui dengan melihat plot *dynamic trace* dan juga dapat dilihat dari nilai *MC error*. *Dynamic trace* merupakan plot nilai dari variabel pada seluruh iterasi yang telah konvergen. Jika *dynamic trace* menunjukkan pola acak (Gambar 2.3) maka iterasi dihentikan dan sebuah contoh acak dikatakan konvergen. *MC error* adalah salah baku dari proses *Markov Chain* (rantai markov) sehingga dapat dikatakan bahwa nilai rata-rata (*mean*) pada metode Bayesian MCMC merupakan koefisien penduga parameter yang terbentuk (Gambar 2.4). Apabila *MC error* bernilai kurang dari 5% simpangan baku maka kekonvergenan dapat terpenuhi dan iterasi dihentikan.

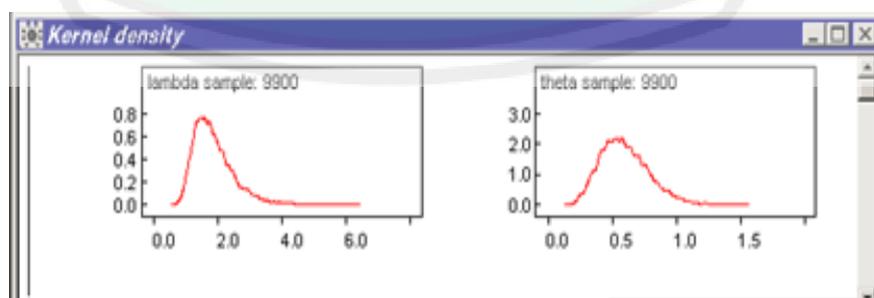


Gambar 2.6 *Dynamic Trace Plot* dari eksponensial parameter

node	mean	sd	MC error 2.5%	median	97.5%	start	sample
lambda	1.862	0.653	0.006134	0.9794	1.734	3.514	101 9900
theta	0.5971	0.1698	0.001813	0.2846	0.5768	1.021	101 9900

Gambar 2.7 penduga parameter sebaran eksponensial

*Kernel density* (Gambar 2.8) digunakan untuk melihat sebaran posterior yang terbentuk dari variabel parameter yang diduga (Tutorial WinBUGS1.4).



Gambar 2.8 *Kernel Density* dari eksponensial parameter.

## 2.12 Kajian *Outlier* dan Estimasi dalam Al-Quran

Al-Quran merupakan firman Allah. Namun Al-Quran bukan hanya berbicara ilmu agama yaitu halal dan haram, pahala dan dosa, surga dan neraka, namun di dalamnya juga terdapat banyak hal yang berkaitan dengan masalah keduniawian, mulai masalah sains dan teknologi, sosial, politik, ekonomi, hukum, dan yang lainnya. Ada banyak sumber kajian tentang itu semua yang menjadikan Al-Quran sebagai acuannya. Oleh sebab itu di sini akan dibuktikan bahwa Al-Quran tidak hanya membahas tentang ilmu agama saja akan tetapi membahas ilmu statistik juga.

Salah satu masalah statistik yang akan dibahas dalam penelitian ini adalah tentang *outlier* dan estimasi. Dalam Al-Quran surat Al-Jin ayat 14, Ar-Ruum ayat 4 dan Qs. Al-Jaatsiah ayat 32.

### 2.12.1 *Outlier* dalam Kajian Al-Quran

Dalam Al-Quran telah disinggung terkait dengan permasalahan *outlier*.

Hal ini terdapat dalam Surat Al-Jin ayat 14:

وَأَنَا مِنَ الْمُسْلِمُونَ وَمِنَّا الْقَاسِطُونَ ۖ فَمَنْ أَسْلَمَ فَأُولَٰئِكَ تَحَرَّوْا رَشَدًا ﴿١٤﴾

Artinya : “Dan sesungguhnya diantara kami ada orang-orang yang taat dan ada (pula) orang-orang yang menyimpang dari kebenaran. Barang siapa yang taat, maka mereka itu benar-benar telah memilih jalan yang lurus”. (Qs. Al-Jin, 72:14).

Asal turunnya surat Al-Jin ayat 14 yaitu untuk menampik dugaan bahwa semua jin baik yang mendengar langsung ayat-ayat Al-Quran maupun yang

belum atau yang tidak mendengarnya kesemuanya telah patuh kepada Allah. Kemudian pada ayat tersebut diterangkan bahwa dan sesungguhnya di antara kami masyarakat jin ada orang-orang muslim yakni yang benar-benar taat dan kepatuhan kepada Allah dan ada pula para penyimpang yakni mereka yang telah sangat jauh dari kebenaran lagi sangat mantap kekufurannya. Barang siapa yang patuh, maka mereka itu telah bersungguh-sungguh memilih arah yang mengantar ke jalan kebenaran (Shihab, 2003:394) .

Dalam surat Al-Jin di atas terdapat kata “*penyimpangan*”. Dalam konsep statistika kata menyimpang diartikan sebagai suatu *outlier*. Sebab suatu *outlier* dikatakan sebagai penyimpang dapat dilihat dari pengertiannya, yaitu:

1. *Outlier* adalah data yang tidak mengikuti pola umum suatu model (Sembiring, 1995:62).
2. *Outlier* adalah suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya (Draper dan Smith, 1992:146).

Dari penafsiran ayat di atas dijelaskan bahwa “para penyimpang yakni mereka yang telah sangat jauh dari kebenaran lagi sangat mantab kekufurannya”. Penafsiran mengenai para penyimpang tersebut mempunyai makna yang sama dengan pengertian dari *outlier* yaitu sama-sama terletak sangat jauh. Namun terdapat perbedaan mengenai konsep *outlier* pada statistika dan pada surat Al-Jin ayat 14 yaitu, dalam statistika suatu data kemungkinan menjadi *outlier* biasanya tidak lebih dari 5 % dari data yang ada. Sedangkan

dalam Surat Al-Jin ayat 14, jumlah penyimpangannya diduga kurang dari 50 % atau bahkan lebih dari 50 %.

### 2.12.2 Estimasi dalam Kajian Al-Quran

Satistika merupakan ilmu yang mempelajari suatu proses dalam mengumpulkan, menganalisis, menginterpretasi dan mempresentasikan data. Sebagian besar konsep dasar statistika adalah mengasumsikan teori probabilitas. Karena statistik bertolak pada cara berfikir probabilistik, hasil pengolahan data yang menggunakan metode statistika bukanlah hasil pasti, melainkan merupakan hasil taksiran adanya ketidakpastian dari variansi yang terjadi dalam fenomena tertentu. Teknik pengambilan kesimpulan tentang suatu parameter meliputi pendugaan parameter dan pengujian hipotesis.

Pendugaan di dalam Al-Quran terdapat dalam penafsiran surat Ar-Ruum ayat 4 :

فِي بَضْعِ سِنِينَ ۗ لِلَّهِ الْأَمْرُ مِنْ قَبْلُ وَمِنْ بَعْدِ ۗ وَيَوْمَئِذٍ يَفْرَحُ الْمُؤْمِنُونَ ﴿٤﴾

Artinya : “Dalam *beberapa tahun lagi*, bagi Allah-lah urusan sebelum dan sesudah (mereka menang). Dan dihari kemenangan bangsa Romawi itu bergembiralah orang-orang yang beriman”.

Dalam Qs. Ar-Ruum ayat 4, terdapat kalimat *فِي بَضْعِ سِنِينَ* (dalam *beberapa tahun lagi*) pengertian lafaz *bid'u sinina* adalah mulai dari tiga tahun sampai dengan sembilan atau sepuluh tahun. Kedua pasukan itu bertemu kembali pada tahun yang ketujuh sesudah pertempuran yang pertama tadi. Akhirnya dalam pertempuran ini pasukan Romawi berhasil mengalahkan

pasukan kerajaan Persia. **لله الامر من قبل ومن بعد** (*bagi Allah-lah urusan sebelum dan sesudahnya*) yakni sebelum bangsa Romawi menang dan sesudahnya. Maksudnya, pada permulaannya pasukan Persia dapat mengalahkan pasukan Romawi, kemudian pasukan Romawi menang atas mereka dengan kehendak Allah. **ويومئذ يفرح المؤمنون** (*bergembiralah orang-orang yang beriman*) (Jalaluddin, 2009: 449).

Dan juga difirmankan pada Qs. Ash-Shaffaat ayat 147, yaitu:

وَأَرْسَلْنَاهُ إِلَىٰ مِائَةِ أَلْفٍ أَوْ يَزِيدُونَ ﴿١٤٧﴾

Artinya: “*dan Kami utus Dia kepada seratus ribu orang atau lebih*”.

Asbabun nuzul pada ayat di atas adalah menceritakan tentang kisah Nabi Yunus. Bahwa ketika Nabi Yunus diancam akan disiksa oleh kaumnya, maka dia keluar dari kalangan mereka sebelum mendapat perintah dari Allah Swt untuk hijrah. Lalu dia naik kapal, namun kapal itu tidak bisa berjalan dan para awak kapal menyangka bahwa kapal itu apabila memuat seorang budak yang melarikan diri, maka kapal itu tidak bisa berjalan. Oleh karena itu mereka melakukan undian dan ternyata undian itu keluar untuk Yunus, maka dilemparkanlah dirinya ke dalam air (Al-Maraghi, 1974:136).

Abdussakir (2007: 155-156) mengatakan bahwa pendugaan (estimasi) adalah keterampilan untuk menentukan sesuatu tanpa melakukan proses perhitungan secara eksak. Disebutkan juga bahwa dalam matematika terdapat

tiga jenis estimasi yaitu estimasi banyak/jumlah (numerositas), estimasi pengukuran, dan estimasi komputasional.

#### 1. Estimasi banyak/ jumlah

Estimasi banyak adalah menentukan banyaknya objek tanpa menghitung secara eksak. Objek disini maknanya sangat luas. Objek dapat bermakna orang, uang, kelereng, titik, dan mobil.

#### 2. Estimasi pengukuran

Estimasi pengukuran adalah menentukan ukuran sesuatu tanpa menghitung secara eksak. Ukuran disini maknanya sangat luas. Ukuran dapat bermakna ukuran waktu, panjang, luas, usia dan volume. Ketika melihat orang berjalan tanpa menanyakan tanggal lahirnya, pembaca dapat menebak/ menaksir usianya. Estimasi pada surat Al-Baqoroh ayat 80 adalah estimasi ukuran yaitu ukuran waktu.

#### 3. Estimasi komputasional

Estimasi komputasional adalah menentukan hasil suatu operasi hitung tanpa menghitungnya secara eksak. Seseorang mungkin akan menghitung dengan cara membulatkan kebuluhan terdekat.

Dari pengertian diatas, maka dapat diketahui kaitan ayat di atas dengan pendugaan adalah terletak dalam kalimat **مائة ألف أو يزيدون**. Karena ayat tersebut di dalam menentukan jumlah umat Nabi Yunus tidak secara perhitungan secara jelas, namun hanya perkiraan saja.

Shihab dalam Tafsir al-Misbah (2003: 84) menjelaskan bahwa kata (أَوْ) pada firmanNya (أَوْ يَزِيدُونَ) oleh sebagian ulama lebih difahami oleh sebagian ulama memahaminya dalam arti *atau* dan ada juga yang memahaminya dalam arti *dan*. Jika dipahami dalam arti *atau*, maka ayat ini bagaikan menyatakan jumlah mereka banyak, yang menurut perhitungannya adalah seratus ribu/ lebih. Dan jika dipahami dalam arti *dan*, maka itu berarti mereka diutus kepada dua kelompok, yang pertama berjumlah seratus ribu (100.000) dan yang satu lagi adalah yang lebih dari itu. Dalam satu riwayat dinyatakan dalam jumlah dua puluh ribu. Yang seratus ribu adalah orang-orang Yahudi penduduk Nainawa, yang ketika itu berada dalam kerajaan Asy'ur, sedang yang lebih adalah selain orang Yahudi yang bermukim juga di negeri itu.

Al-Mahally dan As-Suyuti, dalam Tafsir Jalalain (1990: 640), menjelaskan bahwa وَأَرْسَلْنَاهُ (dan Kami utus dia) sesudah itu, sebagaimana status sebelumnya, kepada kaum Bunainawiy yang tinggal di daerah Mausul. أَوْ أَلْفَ مِائَةٍ (kepada seratus ribu orang atau) bahkan يَزِيدُونَ (lebih dari itu) yakni lebih dua puluh atau tiga puluh atau tujuh puluh ribu orang.

Pendapat yang lain yaitu Amrullah dalam Tafsir Al-Ahzar (1976: 194), menceritakan bahwa setelah Nabi Yunus sehat dan kuat kembali, dia diperintahkan Tuhan melaksanakan perintah yang dipikulny akepadanya, yaitu mendatangi dan melakukan dakwah kepada kaumnya di negeri Ninive ini, yang berjumlah 100.000 orang atau lebih, artinya lebih dari 100.000, kurang tidak. Tugas itupun dilaksanakannya dengan baik karena kesalahan yang telah

diperbuat dahulu itu, lari meninggalkan tugas karena murka/ iba hati kepada kaumnya, telah menginsyafi dan berjanji akan mengubahnya, sebagaimana dalam surat Ash-Shaffat ayat 148:

فَكَامَنُوا.....

Artinya: “*lalu mereka beriman.....*”

Maka berimanlah mereka yaitu kaum Nabi Yunus yang lebih dari seratus itu, merekapun telah beriman.

Para ulama diatas memiliki versi yang berbeda-beda dalam menafsirkan **يزيدون** yang bermakna *lebih*, oleh para ulama diduga sebanyak 20.000 orang, 30.000 orang, atau 70.000 orang. Ada juga ulama yang mengatakan *lebih saja*. Jika dikatakan *lebih saja*, maka bisa saja 10.000 orang atau 15.000 orang. Hal ini disebabkan karena di dalam ayat tersebut tidak dijelaskan secara jelas tentang jumlah umat Nabi Yunus yang sebenarnya. Maka dapat disimpulkan bahwa kata *lebih* disini terdapat batasan tertentu. Jika umat Nabi Yunus dinyatakan dalam matematika adalah  $X$ , maka akan mempunyai interval  $100.000 \leq X \leq 200.000$ , yang artinya umat Nabi Yunus tidak kurang dari 100.000 orang dan tidak akan lebih dari 200.000 orang. Inilah contoh estimasi (taksiran) yang diajarkan Allah kepada kita. Sehingga keterampilan estimasi sangat dibutuhkan dalam kehidupan keseharian kita, karena hal ini sangat menghemat waktu kita dalam sebuah penghitungan.

Berdasarkan penjelasan di atas telah dibuktikan bahwa Al-Quran tidak hanya membicarakan ilmu-ilmu agama saja, akan tetapi juga berbicara tentang

ilmu yang lainnya misalnya saja adalah statistik. Namun, di dalam Al-Quran konsep-konsep ilmu statistik tidak disajikan secara tersirat, akan tetapi berupa pengetahuan yang membutuhkan pengkajian secara mendalam. Itulah sebabnya kenapa Allah SWT memberikan akal dan pikiran kepada manusia. Hal itu disebabkan agar kita mau berpikir dan mengkaji Al-Quran, sehingga kita dapat mengungkap rahasia-rahasia yang terkandung di dalam Al-Quran.



## BAB III

### PEMBAHASAN

Metode bayes merupakan suatu metode untuk menghasilkan estimasi parameter dengan menggabungkan informasi dari sampel dan informasi lain yang telah tersedia sebelumnya. Adapun langkah-langkah dalam mengestimasi model regresi linier dengan pendekatan Bayes adalah sebagai berikut:

#### 3.1 Membentuk Fungsi dari Model Regresi Linier

Bentuk model regresi linier dapat dituliskan sebagai berikut

$$Y = X\beta + \varepsilon \quad (3.1)$$

Keterangan:

$Y$  = vektor data variabel terikat

$X$  = Variabel bebas

$\beta$  = parameter model

$\varepsilon$  = variabel galat, dengan  $\varepsilon \sim N(0, \sigma^2)$

#### 3.2 Membentuk Fungsi Likelihood

Fungsi likelihood diperoleh dengan mengalikan fungsi kepadatan probabilitas  $f(x_i | \theta, \sigma)$  diasumsikan  $e \sim N(0, \sigma^2)$  maka

$$f(e | \sigma) = (2\pi)^{-1/2} (\sigma)^{-1} \exp\left(-\frac{e_i^2}{2\sigma^2}\right) \quad (3.2)$$

$$f(y_i | \theta, \sigma) = (2\pi)^{-1/2} (\sigma)^{-1} \exp\left(-\frac{1}{2\sigma^2} (y_i - x_i\theta)^2\right) \quad (3.3)$$

Untuk membentuk fungsi likelihood  $L(\theta, \sigma | y)$  dari fungsi kepadatan probabilitas  $f(x_i | \theta, \sigma)$  dengan cara

$$L(\theta, \sigma | y) = p(y | \theta, \sigma) = \prod_{i=1}^n f(x_i | \theta, \sigma)$$

Dari persamaan (3.3.2) dapat diperoleh fungsi likelihood

$$\begin{aligned} L(\theta, \sigma | y) &= (2\pi)^{-n/2} (\sigma)^{-n} \exp\left[\left(-\frac{1}{2\sigma^2} (y_1 - x_1\theta)^2 + \dots + (y_n - x_n\theta)^2\right)\right] \\ &= (2\pi)^{-n/2} (\sigma)^{-n} \exp\left[\left(-\frac{1}{2\sigma^2} (y - X\theta)' (y - X\theta)\right)\right] \end{aligned} \quad (3.4)$$

Persamaan (3.4) merupakan fungsi likelihood dari  $f(x_i | \theta, \sigma)$

### 3.3 Menentukan Distribusi Prior Noninformatif

Karena  $e \sim N(0, \sigma^2)$  maka menurut Berger (1985) *prior noninformatif* dari densitas lokasi adalah  $p(\theta)$  dan densitas skala  $p(\sigma)$ . Maka *prior noninformatifnya* diperoleh dari mengalikan densitas lokasi dan densitas skalanya.

#### a. Densitas Lokasi

Misalkan  $\theta$  dan  $\theta^* = \theta + b$  didefinisikan sebagai parameter ruang A, maka dapat diasumsikan suatu *prior* yang memiliki densitas:

$$\begin{aligned} p(\theta \in A) &= p(\theta^* \in A) \\ p(\theta \in A) &= p(\theta + b \in A) \\ p(\theta \in A) &= p(\theta \in A - b) \\ \int_A p(\theta) d\theta &= \int_A p(\theta - b) d\theta \end{aligned}$$

Persamaan tersebut dapat ditulis sebagai  $p(\theta) = p(\theta - b)$ , jika  $\theta = b$  maka  $p(b) = p(\theta)$ . Dimana  $p$  adalah fungsi konstan. Hal ini dapat diasumsikan bahwa  $p = 1$ , sehingga densitas *prior noninformatif* untuk parameter lokasinya adalah:

$$p(\theta) = 1$$

#### b. Densitas Skala

Untuk menentukan densitas skala dimisalkan  $\sigma$  dan  $\sigma^* = d \cdot \sigma$ , yang didefinisikan sebagai parameter dalam ruang  $B$ , maka dapat diasumsikan suatu *prior* yang memiliki densitas

$$p(\sigma \in B) = p(\sigma^* \in B)$$

$$p(\sigma \in B) = p(d\sigma \in B)$$

$$p(\sigma \in B) = p(\sigma \in d^{-1}B)$$

Jika  $\sigma = d$ , maka  $p(d) = \sigma^{-1} p(1)$ .  $p$  adalah fungsi konstan. Hal ini memberikan asumsi bahwa  $p = 1$ , sehingga densitas *prior noninformatif* untuk parameter lokasinya adalah:

$$p(\sigma) = \sigma^{-1}$$

Maka densitas lokasi skala ditentukan dengan persamaan:

$$p(\theta, \sigma) = p(\theta) p(\sigma)$$

$$= 1 \cdot \frac{1}{\sigma}$$

$$= \frac{1}{\sigma}$$

(3.5)

Merupakan prior noninformatif yang memiliki distribusi uniform.

### 3.4 Menentukan *Joint Posterior*

Distribusi joint posterior diperoleh dengan mengalikan fungsi likelihood dan distribusi *priornya*. Sehingga secara umum distribusi posterior diperoleh dari persamaan

$$p(\theta, \sigma | y) \propto L(\theta, \sigma | y) \cdot p(\theta, \sigma)$$

$$\begin{aligned} p(\theta, \sigma | y) &\propto (2\pi)^{-n/2} (\sigma)^{-n} \exp\left[-\frac{1}{2\sigma^2} (y - X\theta)' (y - X\theta)\right] \cdot \frac{1}{\sigma} \\ &\propto (2\pi)^{-n/2} (\sigma)^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2} (y - X\theta)' (y - X\theta)\right] \end{aligned}$$

Karena  $(2\pi)^{-n/2}$  konstanta, maka

$$p(\theta, \sigma | y) \propto (\sigma)^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2} (y - X\theta)' (y - X\theta)\right] \quad (3.6)$$

Misalkan:

$$\begin{aligned} D &= (y - X\theta)' (y - X\theta) \\ &= (y - X\hat{\theta} + X\hat{\theta} - X\theta)' (y - X\hat{\theta} + X\hat{\theta} - X\theta) \\ &= [(y - X\hat{\theta}) - (X\hat{\theta} - X\theta)]' [(y - X\hat{\theta}) - (X\hat{\theta} - X\theta)] \\ &= [(y - X\hat{\theta}) - X(\hat{\theta} - \theta)]' [(y - X\hat{\theta}) - X(\hat{\theta} - \theta)] \\ &= [(y - X\hat{\theta})' - (X(\hat{\theta} - \theta))'] [(y - X\hat{\theta}) - X(\hat{\theta} - \theta)] \\ &= [(y - X\hat{\theta})' - ((\theta - X\hat{\theta})X')] [(y - X\hat{\theta}) - X(\hat{\theta} - \theta)] \\ &= (y - X\hat{\theta})' (y - X\hat{\theta}) + (\hat{\theta} - \theta)' X X' (\hat{\theta} - \theta) - (\hat{\theta} - \theta)' X' (y - X\hat{\theta}) - (y - X\hat{\theta})' X (\hat{\theta} - \theta) \\ &= (y - X\hat{\theta})' (y - X\hat{\theta}) + (\hat{\theta} - \theta)' X X' (\hat{\theta} - \theta) \end{aligned}$$

Maka persamaan (3.6) menjadi

$$p(\theta, \sigma | y) \propto (\sigma)^{-(n+1)} \exp \left[ -\frac{1}{2\sigma^2} (y - \mathbf{X}\hat{\theta})' (y - \mathbf{X}\hat{\theta}) + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right]$$

Selanjutnya persamaan diatas dapat disederhanakan menjadi

$$p(\theta, \sigma | y) \propto \left( \frac{1}{\sigma^{(n+1)}} \right) \exp \left[ -\frac{1}{2\sigma^2} \left( vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right] \quad (3.7)$$

Dimana

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

$$vs^2 = (y - \mathbf{X}\hat{\theta})' (y - \mathbf{X}\hat{\theta})$$

### 3.5 Menentukan Marginal Posterior dari $\theta$

Distribusi Posterior marginal dari  $\theta$  diperoleh dari pengintegralan distribusi posterior terhadap  $\sigma$

$$p(\theta | y) = \int_0^{\infty} p(\theta, \sigma | y) d\sigma$$

Karena  $p(\theta, \sigma | y) \propto (\sigma)^{-n+1} \exp \left[ -\frac{1}{2\sigma^2} \left( vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right]$ , maka

$$p(\theta | y) \propto \int_0^{\infty} \frac{1}{\sigma^{n+1}} \exp \left[ -\frac{1}{2\sigma^2} \left( vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right] d\sigma$$

Misalkan:

$$w = \frac{1}{2\sigma^2} \left[ \left( vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right]$$

$$\frac{dw}{d\sigma} = -\frac{4}{\sigma^3} \left[ \left( vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right]$$

$$d\sigma = \frac{-4\sigma^3}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]} dw$$

Maka,

$$\begin{aligned} p(\theta|y) &\propto \int_0^{\infty} \frac{1}{\sigma^{n+1}} \exp(-w) \frac{-4\sigma^3}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]} dw \\ &\propto \int_0^{\infty} \sigma^{-n+2} \frac{\exp(-w) dw}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]} \\ &\propto \int_0^{\infty} \sigma^{-n+2} \frac{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}}}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}} \frac{2^{\frac{n}{2}-1}}{2^{\frac{n}{2}-1}} \exp(-w) dw} \\ &\propto \int_0^{\infty} \frac{2^{\frac{n}{2}-1}}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}}} \cdot \frac{\sigma^{-n+2} - 1}{2^{\frac{n}{2}-1}} \cdot \left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}-1} \exp(-w) dw \\ &\propto \int_0^{\infty} \frac{2^{\frac{n}{2}-1}}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}}} \cdot \int_0^{\infty} w^{\frac{n}{2}-1} \exp(-w) dw \\ &\propto \frac{2^{\frac{n}{2}-1}}{\left[ \left( v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta}) \right) \right]^{\frac{n}{2}}} \Gamma\left(\frac{n}{2}\right) \end{aligned}$$

Atau

$$p(\theta|y) \propto v s^2 + (\theta - \hat{\theta})' X X (\theta - \hat{\theta})^{-\frac{n}{2}} \quad (3.8)$$

Merupakan fungsi kepadatan probabilitas student-t multivariate.

### 3.6 Menentukan *Marginal Posterior* dari $\sigma$

Distribusi *marginal posterior* dari  $\sigma$  diperoleh dari pengintegralan distribusi *posterior* terhadap  $\theta$ , sehingga

$$p(\sigma | y) = \int p(\theta, \sigma | y) d\theta$$

karena:

$$p(\theta, \sigma | y) \propto (\sigma)^{-(n+1)} \exp \left[ -\frac{1}{2\sigma^2} (y - \mathbf{X}\hat{\theta})' (y - \mathbf{X}\hat{\theta}) + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right]$$

maka distribusi posteriornya di dapat

$$p(\sigma | y) \propto \int_{-\infty}^{\infty} (\sigma)^{-(n+1)} \exp \left[ -\frac{1}{2\sigma^2} \left( v s^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right) \right] d\theta$$

Karena diintegrasikan terhadap  $\theta$ , maka  $\frac{1}{\sigma^{-(n+1)}} \exp \left( -\frac{1}{2\sigma^2} v s^2 \right)$  dapat

dianggap sebagai konstanta, sehingga

$$p(\sigma | y) \propto \frac{1}{\sigma^{n+1}} \exp \left( \frac{v s^2}{2\sigma^2} \right) \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right] d\theta$$

Dimisalkan:

$$\int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2\sigma^2} (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X} (\theta - \hat{\theta}) \right] d\theta = I$$

$Z = \frac{1}{\sigma} (\theta - \hat{\theta})$ , maka  $\theta = \sigma Z + \hat{\theta}$

dan Jacobian transformasi  $J = \sigma^p$ , sehingga

$$\begin{aligned}
 I &= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} Z' X X Z\right] \sigma^p \partial Z \\
 &= \sigma^p \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} Z' X X Z\right] \partial Z \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|} \times \frac{1}{\sqrt{(2\pi)^p |(X X)^{-1}|}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} Z' X X Z\right] \partial Z \\
 &= \sigma^p \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} Z' X X Z\right] \partial Z \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|} \times \frac{1}{\sqrt{(2\pi)^p |(X X)^{-1}|}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} Z' X X Z\right] \partial Z \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|} \times 1 \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|} \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|} \times 1 \\
 &= \sigma^p \sqrt{(2\pi)^p |(X X)^{-1}|}
 \end{aligned}$$

Dan karena hasil integralnya

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2} (\theta - \hat{\theta})' X X (\theta - \hat{\theta})\right] \partial \theta = \sigma^p (2\pi)^{p/2} |(X X)^{-1}|^{1/2} \text{ maka}$$

$$p(\sigma | y) \propto \frac{1}{\sigma^{n+1}} \exp\left(-\frac{v s^2}{2\sigma^2}\right) \sigma^p (2\pi)^{p/2} |(X X)^{-1}|^{1/2}$$

$$\propto \frac{1}{\sigma^{n+1}} \exp\left(-\frac{v s^2}{2\sigma^2}\right) \frac{1}{\sigma^{-p}} (2\pi)^{p/2} |(X X)^{-1}|^{1/2}$$

(3.9)

karena  $(2\pi)^{p/2} |(\mathbf{X}'\mathbf{X})^{-1}|^{1/2}$  dianggap konstanta maka,

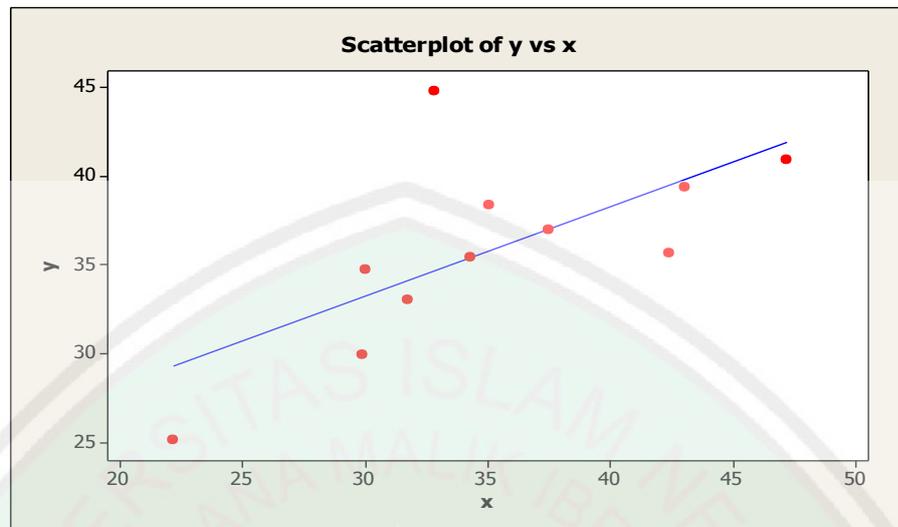
$$\begin{aligned}
 p(\sigma | y) &\propto \frac{1}{\sigma^{n-p+1}} \exp\left(-\frac{vs^2}{2\pi^2}\right) \\
 &\propto \frac{1}{\sigma^{v+1}} \exp\left(-\frac{vs^2}{2\pi^2}\right)
 \end{aligned}
 \tag{4.10}$$

### 3.7 Penerapan pada Studi Kasus Data Curah Hujan di Seattle dan Portland yang Disimulasikan dengan Menggunakan Bantuan Paket Program *Windows Bayesian Inference Using Gibbs Sampling (WinBUGS)* versi 1.4.

Data berikut berasal dari Birkes dan Dodge (1993)

Tabel 3.1 Data Curah Hujan tahunan (dalam inci) di Seattle dan di Portland

Tahun	Curah Hujan di Seattle (y)	Curah Hujan di Portland (x)
1980	35.60	42.41
1981	35.40	34.29
1982	39.32	43.04
1983	40.93	47.19
1984	36.99	37.50
1985	25.13	22.18
1986	38.34	35.04
1987	29.93	29.91
1988	32.98	31.72
1989	34.69	30.05
1990	44.75	32.86



Gambar 3.1 Sebaran data yang mengandung outlier

Dari gambar diatas terdapat data yang menyimpang dari data lain yaitu pada  $(32.86, 44.75)$  dan  $(22.18, 25.13)$  yang biasanya disebut dengan *outlier*. Data tersebut bisa saja berpengaruh terhadap penduga parameter regresinya, salah satu cara mengatasi masalah ini adalah dengan memeriksa data tersebut yaitu dengan analisis regresi yang terdapat *outlier*.

Adapun langkah-langkah penerapannya adalah:

### 3.7.1 Melakukan Pengecekan Data Apakah Terdapat *Outlier* atau Tidak

Tabel 3.2 Data yang terdapat *outlier*

Tahun	Curah Hujan di Seattle (y)	Curah Hujan di Portland (x)
1980	35.60	42.41
1981	35.40	34.29
1982	39.32	43.04
1983	40.93	47.19
1984	36.99	37.50
1985	25.13	22.18
1986	38.34	35.04
1987	29.93	29.91
1988	32.98	31.72
1989	34.69	30.05
1990	44.75	32.86

Adapun model regresi dari data di atas adalah

$$y_i = \beta_0 + \beta_1 x_n + \varepsilon_i$$

Sehingga bentuk dalam matriksnya adalah

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{(11)t} \end{pmatrix} = \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{11} \end{pmatrix} = \begin{pmatrix} 35.60 \\ 35.40 \\ \vdots \\ 44.75 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{11} \end{pmatrix}$$

Dan untuk menentukan parameter  $\boldsymbol{\beta}$  menggunakan rumus penduga parameter

$\hat{\boldsymbol{\beta}}$  yang diperoleh dari metode *Maximum Likelihood* yaitu

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Sehingga

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{bmatrix} \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix}' \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix} \end{bmatrix}^{-1} \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix} \begin{pmatrix} 35.60 \\ 35.40 \\ \vdots \\ 44.75 \end{pmatrix}$$

$$= \begin{bmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 42.41 & 34.29 & \dots & 32.86 \end{pmatrix} \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix} \end{bmatrix}^{-1} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 42.41 & 34.29 & \dots & 32.86 \end{pmatrix} \begin{pmatrix} 35.60 \\ 35.40 \\ \vdots \\ 44.75 \end{pmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} 11 & 386 \\ 386 & 14063 \end{bmatrix}^{-1} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 42.41 & 34.29 & \dots & 32.86 \end{pmatrix} \begin{pmatrix} 35.60 \\ 35.40 \\ \vdots \\ 44.75 \end{pmatrix} \\
&= \frac{1}{(11 \times 14063) - (386 \times 386)} \begin{bmatrix} 14063 & -386 \\ -386 & 11 \end{bmatrix} \begin{bmatrix} 394 \\ 14090 \end{bmatrix} \\
&= \frac{1}{(154693) - (148996)} \begin{bmatrix} 14063 & -386 \\ -386 & 11 \end{bmatrix} \begin{bmatrix} 394 \\ 14090 \end{bmatrix} \\
&= \frac{1}{(154693) - (148996)} \begin{bmatrix} 14063 & -386 \\ -386 & 11 \end{bmatrix} \begin{bmatrix} 394 \\ 14090 \end{bmatrix} \\
&= \frac{1}{5697} \begin{bmatrix} 14063 & -386 \\ -386 & 11 \end{bmatrix} \begin{bmatrix} 394 \\ 14090 \end{bmatrix} \\
&= \begin{bmatrix} 18.0952 \\ 0.5050 \end{bmatrix}
\end{aligned}$$

Maka

$$\hat{Y} = X \hat{\beta} = \begin{pmatrix} 1 & 42.41 \\ 1 & 34.29 \\ \vdots & \vdots \\ 1 & 32.86 \end{pmatrix} \begin{bmatrix} 18.0952 \\ 0.5050 \end{bmatrix} - 0.0105$$

Tabel 3.3 Analisis data regresi yang mengandung *outlier*

No	$x$	$y$	$\hat{y} = x\hat{\beta}$	$\varepsilon = y - \hat{y}$
1	42.41	35.60	39.5108	-3.9108
2	34.29	35.40	35.4105	-0.0105
3	43.04	39.32	39.8289	-0.5089
4	47.19	40.93	41.9246	-0.9946
5	37.50	36.99	37.0314	-0.0414
6	22.18	25.13	29.2953	-4.1653
7	35.04	38.34	35.7892	2.5508
8	29.91	29.93	33.1987	-3.2687
9	31.72	32.98	34.1127	-1.1327
10	30.05	34.69	33.2694	1.4206
11	32.86	44.75	34.6884	10.0616

Penyelesaian untuk estimasi marginal posterior setiap parameter model akan didekati dengan metode *Markov Chain Monte Carlo*. Berdasarkan *marginal posterior* yang diperoleh di atas untuk setiap parameter dalam model regresi, maka estimasi parameter model dalam MCMC akan cukup berbeda caranya. Hasil *joint posterior* di atas digunakan sebagai awalan dalam membangun MCMC untuk model regresi ini. MCMC khususnya *Gibbs Sampler* akan menirukan proses markov, dimana mencatat proses sekarang dipengaruhi oleh satu step sebelumnya. Sehingga dalam proses *Gibbs Sampler* akan membangun step-step proses MCMC-nya dengan membuat *full conditional posterior* dan menyusunnya bergantian sebagai step iteratif simulasi stokhastiknya.

### 3.7.2 Menduga Parameter dengan Menggunakan *Ordinary Least Square* (OLS) dan Bayesian MCMC

Setelah diketahui bahwa data tersebut terdapat *outlier*, maka langkah selanjutnya adalah dilakukan pendugaan parameter dengan menggunakan *Ordinary Least Square* (OLS) dan pendugaan parameter dengan menggunakan Bayesian MCMC. Hal ini bertujuan untuk mengetahui hasil manakah yang menunjukkan model terbaik antara pendugaan model dengan menggunakan metode klasik dan metode Bayesian MCMC.

#### 3.7.2.1 Pendugaan Parameter dengan *Ordinary Least Square* (OLS)

Pada subbab ini, akan disajikan hasil pendugaan parameter dengan *Ordinary Least Square* (OLS) yang dalam analisisnya menggunakan MINITAB. Adapun hasil pendugaan parameter menggunakan OLS disajikan dalam tabel 4.4 dan hasil selengkapnya dapat dilihat pada lampiran 1.

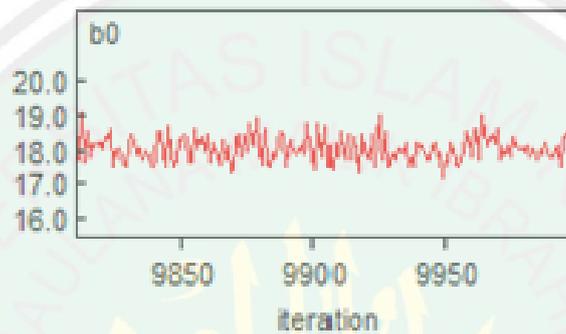
Tabel 3.4 Hasil Pendugaan Parameter dengan menggunakan MINITAB

Parameter	Koefisien	Simpangan baku
$\beta_0$	18.1	6.617
$\beta_1$	0.505	0.1851

#### 3.7.2.2 Pendugaan Parameter dengan Bayes MCMC

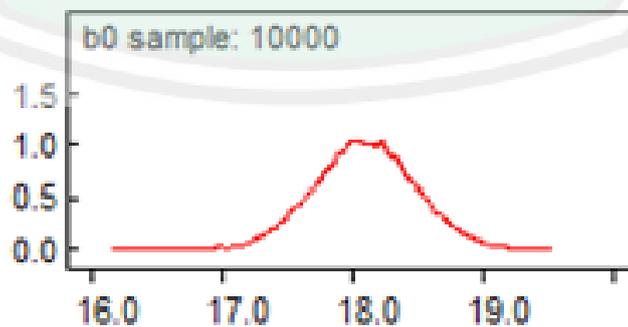
Data pada penelitian ini digunakan sebagai nilai awal dalam proses simulasi metode Bayes MCMC. Proses simulasi diawali dengan membentuk model berdasarkan Model *Specification* (program dapat dilihat pada lampiran 2), dan kemudian dilakukan iterasi dengan *Gibbs Sampler*. Iterasi dilakukan

sebanyak 10000 sampel . Sebuah contoh acak dikatakan konvergen apabila tidak membentuk sebuah pola yang teratur. Dimana di dalam WinBugs 1.4 dapat kita lihat dalam grafik *dynamic trace*. Salah satu *dynamic trace* dari model penelitian ini adalah



Gambar 3.2 *Dynamic Trace*

Pada gambar 3.2 tersebut terlihat bahwa tidak membentuk sebuah pola yang teratur, sehingga dapat dikatakan bahwa contoh acak tersebut telah konvergen. Sebaran posterior yang terbentuk dari hasil penelitian ini dapat dilihat dari *kernel density*, adapun hasil output WinBugs 1.4, salah satu sebaran posterior yang dihasilkan adalah:



Gambar 3.3 *Kernel Density*

Pada gambar 3.3 tersebut terlihat bahwa sebaran posterior yang terbentuk untuk parameter  $\beta_0$  berbentuk hampir menyerupai sebaran normal. Adapun hasil pendugaan parameter dengan menggunakan pendugaan Bayes MCMC terlihat dalam Tabel 3.5.

Tabel 3.5 Hasil Pendugaan Parameter dengan menggunakan WinBUGS

Parameter	Rata-rata	Simpangan baku	MC error
$\beta_0$	18.09	0.3894	0.003893
$\beta_1$	0.5047	0.02638	$2.769E-4$

Pada tabel 3.5 menunjukkan bahwa rata-rata adalah rata-rata dari beberapa sampel iterasi pada masing-masing parameter. Nilai tersebut digunakan sebagai penduga parameter dalam model regresi linier sederhana. Sedangkan simpangan baku yang dihasilkan merupakan simpangan baku dari beberapa sampel iterasi dan MC *error* adalah simpangan baku dari proses *Markov Chain* (rantai markov).

### 3.7.3 Analisis Perbandingan Model

Di dalam melakukan perbandingan model regresi yang dibangun dengan OLS dan Bayesian MCMC, maka dalam analisisnya dipilih OLS dengan menggunakan MINITAB dan Bayes MCMC dengan menggunakan WinBUGS. Adapun hasil olahan untuk masing-masing metode dapat dilihat dalam dua gambar keluaran program MINITAB dan WinBUGS yang terdapat dalam lampiran 2 dan lampiran 5.

Dari dua hasil pemodelan yang menggunakan OLS dan Bayes yaitu dengan membandingkan nilai taksiran varian model, maka dapat dikatakan

bahwa dengan model Bayesian MCMC yang diperoleh menunjukkan lebih baik dibandingkan dengan *Ordinary Least Square* (OLS). Hal ini disebabkan karena *Mean Square Error* (MSE) dalam OLS jauh lebih besar yaitu 17.29 daripada *Mean Square Error* (MSE) dalam Bayes MCMC yaitu 8.316.

### 3.8 Korelasi Al Quran dengan Matematika

Matematika oleh sebagian orang lebih banyak dikenal sebagai disiplin ilmu yang tidak memiliki kaitan dengan keislaman banyak pendapat yang mengatakan bahwasannya matematika merupakan ilmu yang dihasilkan oleh orang-orang Barat sehingga di dalam kajiannya jauh dari nilai-nilai spiritual. Bahkan ada juga pihak instansi pendidikan “Islam” yang tidak membolehkan matematika untuk diajarkan kepada anak didiknya. Inilah sekilas fakta yang masih menjangkit masyarakat di sekitar kita.

Sesungguhnya matematika itu memiliki hubungan yang sangat erat dengan tradisi spiritual umat Islam, dan matematika juga dapat dijadikan “jalan” menuju pencapaian manfaat kebahagiaan baik di dunia maupun di akhirat. Matematika berada pada posisi di antara dunia nyata dan dunia ghaib. Matematika tidak berada dalam dunia nyata sehingga objek yang dikajian bersifat abstrak dan tidak berada di dunia ghaib sehingga objek matematika bukanlah sesuatu “penampakan”. Membawa objek dunia nyata ke dalam bahasa matematika disebut dengan abstrak dan mewujudkan matematika dalam dunia nyata disebut dengan aplikasi.

Matematika berada dalam dunia diantara dunia syahadah dan ghaibiyah. Dengan demikian, maka matematika bersifat “setengah nyata dan setengah ghaib”. Untuk memahami objek yang nyata diperlukan pendekatan rasionalis, empiris, dan logis (bayani dan burhani). Sedangkan untuk memahami objek yang ghaib diperlukan pendekatan intuitif, imajinatif. Dan metafisis (irfani). Kekuatan utama dalam matematika terletak pada imajinasi atau intuisi yang kemudian diterima setelah dibuktikan secara logis atau deduktif. Dengan demikian, maka untuk mempelajari matematika perlu penggabungan ketiga pendekatan tersebut, yaitu bayani, burhani, dan ‘irfani.

Dalam kajian penelitian ini, penulis menguraikan tentang aspek-aspek matematika yang tersirat dalam Al-Quran. Penulis ingin membuktikan bahwa ternyata di dalam Al-Quran itu juga membicarakan konsep matematika. Hal ini akan dapat mematahkan “kepercayaan” sebagian orang yang meyakini bahwa matematika itu produk Barat. Berbagai macam ilmu yang kita pelajari selama ini tidaklah terlepas dari Al-Quran. Sebagaimana dalam kajian Bab II, kajian *outlier* terdapat dalam Qs. Al-Jin ayat 14 dan dijelaskan pula pendugaan parameter yang terdapat dalam Al-Quran Surat Ash-Shaffat ayat 147. Dalam bab ini, akan diuraikan tentang korelasi antara Qs. Al-Jin ayat 14 dengan konsep *outlier* dalam matematika dan Qs. Ash-Shaffaat Ayat 147 dengan konsep pendugaan dalam matematika.

Konsep *outlier* dalam matematika adalah data yang menyimpang yang tidak mengikuti pola umum suatu model. Pada Qs. Al-Jin, 72:14 dijelaskan bahwasannya dalam suatu kaum jin itu terdiri dari dua macam, yaitu kaum

yang taat kepada Allah (المسلمون) dan kaum yang menyimpang dari kebenaran (القسطون). Di dalam matematika kaum yang menyimpang dari kebenaran Allah itulah yang didefinisikan sebagai *outlier*. Apabila dihubungkan dalam konsep matematika, kaum yang menyimpang tersebut letaknya sangat jauh dari kebenaran bahkan sampai pada tingkat kekufuran. Yang dapat digambarkan dalam grafik



Gambar 3.4 Menunjukkan tiga titik penyimpangan

Pada gambar 4.4 di atas dapat dilihat bahwa terdapat tiga titik data yang terpisah sangat jauh dari data lainnya yaitu data 1, 2, dan 3. Ketiga titik itulah yang disebut sebagai *outlier* yaitu umat kaum jin yang menyimpang dari kebenaran Allah dan bahkan benar-benar kufur kepada Allah.

Namun terdapat perbedaan mengenai konsep *outlier* pada statistik dan pada Qs. Al-Jin yaitu, dalam statistik suatu data kemungkinan menjadi *outlier* biasanya tidak lebih dari 5% dari data yang ada, sedangkan dalam Qs.Al-Jin, jumlah penyimpangannya diduga kurang dari 50% atau bahkan lebih dari 50% .

Konsep pendugaan dalam matematika ternyata telah terkonsep sejak zaman Nabi Muhammad saw. Hal tersebut terbukti dalam Al-Quran Surat Ash-Shaffat ayat 147, yang secara tidak tersirat telah mengkaji tentang konsep pendugaan.

وَأَرْسَلْنَاهُ إِلَىٰ مِائَةِ أَلْفٍ أَوْ يَزِيدُونَ ﴿١٤٧﴾

Artinya: “ dan Kami utus Dia kepada seratus ribu orang atau lebih”.

Penafsiran pendugaan dalam Surat Ash-Shaffat ayat 147 merupakan perkiraan, maksudnya adalah menghitung jumlah umat Nabi Yunus tidak secara eksak, yaitu melalui penaksiran atau memperkira-kirakan saja. Dari sini dapat diketahui bahwasannya pendugaan dalam ayat tersebut merupakan pendugaan dalam konsep yang sederhana dan dalam matematika digunakan untuk perhitungan-perhitungan dasar matematika. Dengan seiring berkembangnya zaman, berkembang pula ilmu pengetahuan. Konsep pendugaan dalam Surat Ash-Shaffat ayat 147 merupakan konsep dasar matematika yang kemudian dikembangkan salah satunya dalam bidang statistika, adapun pengertian pendugaan dalam statistik adalah proses yang menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi yang tidak diketahui.

Perbedaan pendugaan dalam Surat Ash-Shaffat dengan pendugaan parameter dalam penelitian ini terletak pada objeknya. Dalam Surat Ash-Shaffat menduga terhadap banyaknya jumlah dan syarat penduga berupa interval yaitu  $100.000 \leq X \leq 200.000$ , sedangkan dalam penelitian ini

menduga model regresi yang pendugaannya berupa rumus yang dapat diterapkan dalam penelitian-penelitian lapangan.

Dari sinilah perlu diketahui, bahwa ilmu pengetahuan umum seperti matematika khususnya konsep pendugaan parameter yang diciptakan oleh orang-orang barat nonmuslim, ternyata telah terkonsep dalam Al-Quran. Hal ini membuktikan bahwa Al-Quran tidak hanya berbicara tentang halal dan haram, serta ilmu-ilmu agama saja melainkan juga berbicara tentang ilmu pengetahuan umum. Namun dalam Al-Quran, konsep-konsep ilmu pengetahuan umum tidak dijabarkan secara langsung, akan tetapi pengetahuan yang membutuhkan penafsiran secara mendalam. Oleh karena itu Allah SWT memberikan akal kepada manusia, agar supaya manusia bisa berpikir dan mengkaji Al-Quran sehingga bisa menguak rahasia-rahasia yang terkandung dalam Al-Quran.

Matematika dapat dijadikan sumber pelajaran dalam rangka menapaki hidup menuju ridha-Nya. Dengan upaya pemaknaan secara “Islami” inilah diharapkan dapat mengobati “luka” lama umat Islam terhadap “sakit apatisme” pada matematika yang selama ini telah menjangkitnya. Akibatnya akan muncul gerakan “sadar matematika” di dunia ini, sehingga kejayaan dan peradaban Islam akan dapat dicapai kembali. Sudah saatnya sekarang ini umat Islam mampu berkompetisi secara sehat dalam persaingan di dunia global. Sudah tidak ada lagi alasan untuk menolak kehadiran kecanggihan informasi dan teknologi. Tetapi, salah satu yang perlu diingat bahwa semua yang kita tekuni

harus tetap bersumber pada landasan Al-Quran dan Al-Hadist. Maka kebahagiaan dunia akhirat dapat kita capai secara bersama-sama.



## BAB IV

### PENUTUP

#### 4.1 KESIMPULAN

1. Desain pemodelan Bayes MCMC dengan menggunakan WinBUGS mempunyai basis pemodelan grafik lebih memudahkan untuk mengimplementasikan proses simulasi stokhastik sebuah pemodelan regresi linier
2. Dengan menggunakan distribusi prior hasil pemodelan analitik yang dipadukan dengan parameter setiap diatribusi prior dan hasil *Ordinary Least Square* (OLS) akan dapat diperoleh hasil estimasi Bayes MCMC yang lebih baik dari pada *Ordinary Least Square* (OLS).
3. Hasil dari penelitian ini didapatkan joint posterior 
$$\left(\frac{1}{\sigma^{(n+1)}}\right) \exp\left[-\frac{1}{2\sigma^2}\left(vs^2 + (\theta - \hat{\theta})' \mathbf{X}'\mathbf{X}(\theta - \hat{\theta})\right)\right]$$
. Hasil joint posterior tersebut akan digunakan sebagai awalan dalam membangun MCMC untuk model regresinya. MCMC khususnya *Gibbs Sampler* yang digunakan disini akan menirukan proses Markov yang mencatat proses sekarang akan dipengaruhi satu step proses sebelumnya.
4. Dari hasil pemodelan tersebut menunjukkan metode Bayes MCMC lebih baik dibandingkan *Ordinary Least Square* (OLS). Hal ini disebabkan karena *Mean Square Error* (MSE) dalam OLS jauh lebih besar yaitu 17.29 daripada *Mean Square Error* (MSE) dalam Bayes MCMC yaitu 8.316.

## 4.2 SARAN

Saran yang dapat diberikan untuk penelitian selanjutnya adalah diharapkan menggunakan teknik *Metropolis Hasting* untuk membangkitkan peubah acak dari sebaran tertentu dalam metode *Markov Chain Monte Carlo* yang cara kerjanya lebih sederhana dari pada teknik simulasi *Gibbs Sampling*.



## DAFTAR PUSTAKA

- Abdussakir. 2007. *Ketika Kyai Mengajar Matematika*. UIN-Malang press: Malang.
- Aziz, Abdul. 2007. *Ekonometrika, Teori Analisis Matematika dilengkapi Eksperimen dengan Matlab*. Jakarta: Prestasi Pelajar.
- Box, George E.P and Tiao, George C. 1973. *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley Publishing Company.
- Draper, Norman dan Smith, Harry. 1992. *Analisis Regresi Terapan Edisi Kedua*. Jakarta: PT. Gramedia Pustaka Utama.
- Firdaus, Muhammad. 2004. *Ekonometri Suatu Pendekatan Aplikatif*. Jakarta: PT Bumi Aksara.
- Hasan, M. Iqbal. 2002. *Pokok-Pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta : PT Bumi Aksara.
- Jalaluddin, Al-Mahalli, Imam dan Jalaluddin As-Suyuti, Imam. 2009. *Tafsir Jalalain*. Bandung: Sinar Baru Algensindo.
- Lains, Alfian. 2003. *Ekonometrika Teori dan Aplikasi*. Jakarta: Pustaka LP3ES Indonesia.
- Pereira, F. 1999. *Practical Modern Bayesian Statistics In Actuarial Science*. General Insurance Convention.
- Scollnik, D.P.M. 1996. *An Intproduction To Markov Chain Monte Carlo and Their Actuarial Applications. Proceedind The Casuality Society*. Department of Mathematics and Statistics. University of Calgary.
- Seber, George A.F and Lee, Alan J. 2003. *Linear Regression Analysis*. Canada: Wiley interscience.
- Sembiring. 1995. *Analisis Regresi*. Bandung : ITB.
- Steel, Robert G.D. and Torri, James H. 1989. *Prinsip dan Prosedur Statistika Suatu Pendekatan Biometrik*. Jakarta: Gramedia.
- Supranto. 2004. *Ekonometri*. Jakarta: Ghalia Indonesia.

Turmudi, dan Harini, Sri. 2008. *Metode Statistika: Pendekatan Teoritis dan Aplikatif*. Malang: Uin-Malang Press.

Walpole, Ronald E. 1995. *Pengantar Statistika Edisi Ketiga*. Jakarta: PT Gramedia Pustaka Utama.

Yitnosumarto, Suntoyo. 1990. *Dasar-Dasar Statistika*. Jakarta: C.V Rajawali.



**Lampiran 1. Data Curah Hujan di Seattle dan Portland**

Tahun	Curah Hujan di Seattle ( $y$ )	Curah Hujan di Portland ( $x$ )
1980	35.60	42.41
1981	35.40	34.29
1982	39.32	43.04
1983	40.93	47.19
1984	36.99	37.50
1985	25.13	22.18
1986	38.34	35.04
1987	29.93	29.91
1988	32.98	31.72
1989	34.69	30.05
1990	44.75	32.86

Keterangan:

$y$  = Curah Hujan di Seattle

$x$  = Curah Hujan di Portland

**Lampiran 2. Hasil Analisis OLS pada Data Curah Hujan dengan menggunakan MINITAB**

**Regression Analysis: Curah Hujan di S versus Curah Hujan di P**

The regression equation is  
 Curah Hujan di Seattle (y) = 18,1 + 0,505 Curah Hujan di Portland (x)

Predictor	Coef	SE Coef	T	P
Constant	18,095	6,617	2,73	0,023
Curah Hujan di Portland (x)	0,5050	0,1851	2,73	0,023

S = 4,15830    R-Sq = 45,3%    R-Sq(adj) = 39,2%

Analysis of Variance

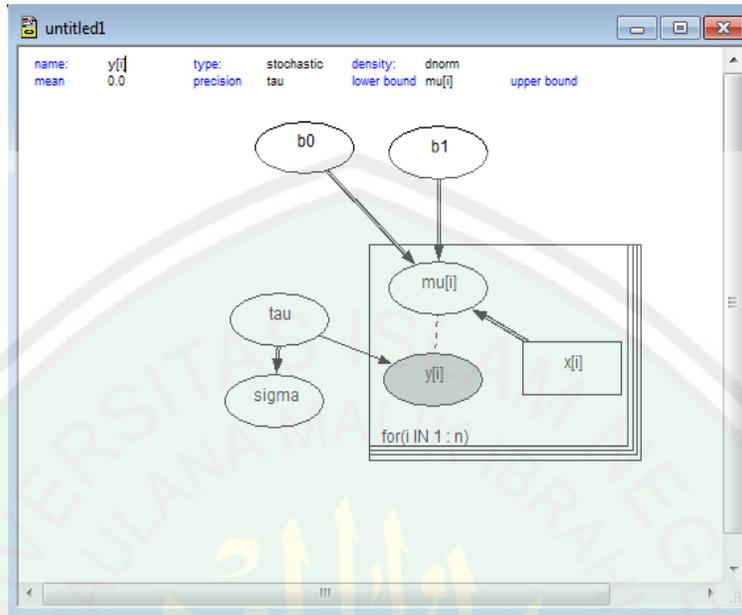
Source	DF	SS	MS	F	P
Regression	1	128,74	128,74	7,45	0,023
Residual Error	9	155,62	17,29		
Total	10	284,36			

Unusual Observations

Obs	Curah Hujan di Portland (x)	Curah Hujan di Seattle (y)	Fit	SE Fit	Residual	St Resid
11	32,9	44,75	34,69	1,32	10,06	2,55R

R denotes an observation with a large standardized residual.

### Lampiran 3. Struktur Doodle dalam WinBugs



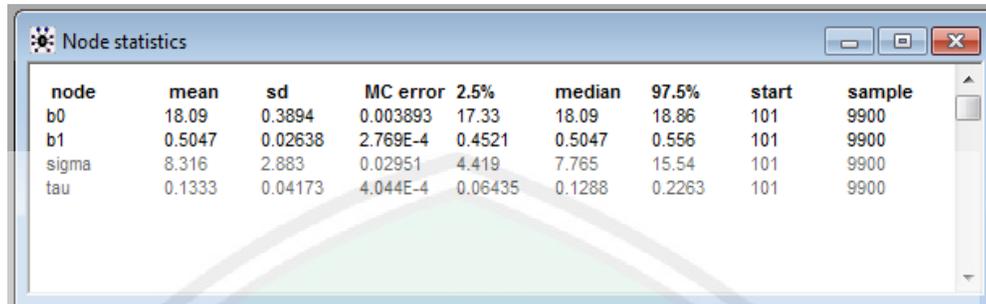
#### Lampiran 4. Program WinBugs 1.4 untuk data Penelitian

```
model;
{
  for(i in 1 : n) {
    y[i] ~ dnorm(mu[i],tau)
  }
  tau ~ dgamma( 5.5,1)
  sigma <- 1 / tau
  b1 ~ dnorm(0.505,0.1851)
  b0 ~ dnorm(18.09,6.617)
  for(i in 1 : n) {
    mu[i] <- b0 + b1 * x[i]
  }
}

inisialisasi
list(b0=0,b1=0,tau=0.1)

data
list(n=11,
y=c(35.60,35.40,39.32,40.93,36.99,25.13,38.34,29.93,32.98,34.69,44.75),
x=c(42.41,34.29,43.04,47.19,37.50,22.18,35.04,29.91,31.72,30.05,32.86))
```

Lampiran 5. Hasil Output Program WinBugs 1.4 pada data penelitian

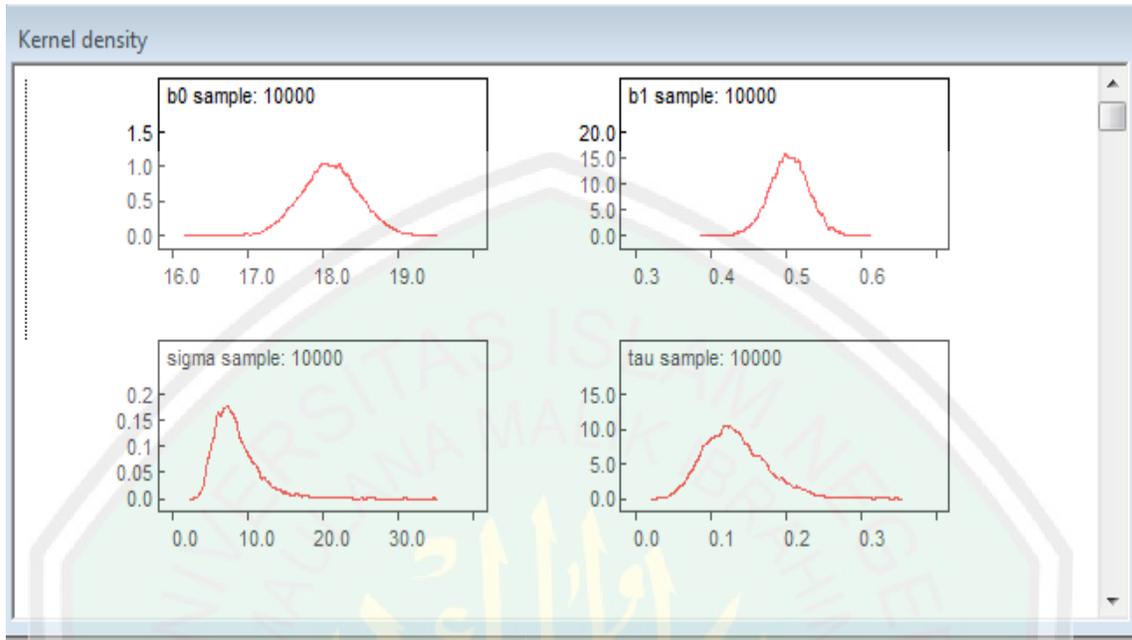


node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b0	18.09	0.3894	0.003893	17.33	18.09	18.86	101	9900
b1	0.5047	0.02638	2.769E-4	0.4521	0.5047	0.556	101	9900
sigma	8.316	2.883	0.02951	4.419	7.765	15.54	101	9900
tau	0.1333	0.04173	4.044E-4	0.06435	0.1288	0.2263	101	9900



Lampiran 6.

Gambar Kernel Dencity penduga parameter



Gambar Dynamic Trace penduga parameter

