ESTIMASI PARAMETER MODEL REGRESI LINIER PADA DATA YANG MENGANDUNG *OUTLIER* DENGAN METODE *MAXIMUM LIKELIHOOD ESTIMATION*

SKRIPSI

Oleh:
NUR NGAINI
NIM. 08610072

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2012

ESTIMASI PARAMETER MODEL REGRESI LINIER PADA DATA YANG MENGANDUNG *OUTLIER* DENGAN METODE *MAXIMUM LIKELIHOOD ESTIMATION*

SKRIPSI

Diajukan Kepada:
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Sains (S.Si)

Oleh:
NUR NGAINI
NIM. 08610072

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2012

ESTIMASI PARAMETER MODEL REGRESI LINIER PADA DATA YANG MENGANDUNG *OUTLIER* DENGAN METODE *MAXIMUM LIKELIHOOD ESTIMATION*

SKRIPSI

Oleh:
NUR NGAINI
NIM. 08610072

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 12 Januari 2012

Pembimbing I	**Pembimbing II**
Sri Harini, M.Si	**Dr. Ahmad Barizi, MA**
NIP. 19731014 200112 2 002	**NIP. 19731212 199803 1 001**

Mengetahui,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP. 19751006 200312 1 001

ESTIMASI PARAMETER MODEL REGRESI LINIER PADA DATA YANG MENGANDUNG *OUTLIER* DENGAN METODE *MAXIMUM LIKELIHOOD ESTIMATION*

SKRIPSI

Oleh:
NUR NGAINI
NIM. 08610072

Telah Dipertahankan di Depan Dewan Penguji Skripsi dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar Sarjana Sains (S.Si)
Tanggal: 21 Januari 2012

Penguji Utama	: **Drs. Turmudi, M.Si** **NIP. 19571005 198203 1 006**	…………………….
Ketua Penguji	: **Abdul Aziz, M.Si** **NIP. 19760218 200604 1 002**	…………………….
Sekretaris Penguji	: **Sri Harini, M.Si** **NIP. 19731014 200112 2 002**	…………………….
Anggota Penguji	: **Dr. Ahmad Barizi, MA** **NIP. 19731212 199803 1 001**	…………………….

Mengesahkan,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP. 19751006 200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama	: Nur Ngaini
NIM	: 08610072
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. Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 10 Januari 2012
Yang membuat pernyataan,

Nur Ngaini
NIM. 08610072

MOTTO

أَلَا بِذِكۡرِ ٱللَّهِ تَطۡمَئِنُّ ٱلۡقُلُوبُ

"Ingatlah, Hanya dengan mengingati Allah-lah

hati menjadi tenteram (Qs. Ar-Ra'd:28)"

*Ketika hidup memberi kata "**tidak**"*

Atas apa yang kamu inginkan

*Percayalah Allah selalu memberi kata "**ya**"*

Atas apa yang kamu butuhkan

PERSEMBAHAN

Karya ini penulis persembahkan untuk. . . . . .

Ayah dan Ibu

Ya Allah….melalui merekalah

Engkau ajarkan setiap kata, setiap senyuman, ketabahan, kesabaran dan arti sebuah kehidupan.

Engkau ajarkan kasih sayang yang tak pernah ada henti-hentinya sampai kapanpun.

Kasih sayang yang tak senilai dan tak sebanding dengan sesuatu apapun yang dapat membalasnya.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Syukur alhamdulillah penulis hanturkan ke hadirat Allah SWT yang telah melimpahkan rahmat, taufik serta hidayah-Nya sehingga penulisan skripsi yang berjudul **"Estimasi Parameter Model Regresi Linier pada Data yang Mengandung *Outlier* dengan Metode *Maximum Likelihood Estimation*"** dapat terselesaikan dengan baik, sekaligus dapat menyelesaikan studi di Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Sholawat serta salam semoga tetap tercurah dan terlimpahkan kepada Nabi Muhammad SAW, yang telah membawa manusia dari zaman jahiliyah menuju zaman yang terang penuh cahaya keimanan yakni Ad-dinul Islam Wa Nurul Iman.

Penulis menyadari bahwa skripsi ini tidak akan selesai dengan baik tanpa adanya saran, arahan, bimbingan, serta do'a dan bantuan dari semua pihak. Oleh karena itu dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Prof. Dr. H. Imam Suprayogo, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Drs. Sutiman Bambang Sumitro, S.U, 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 dan Ahmad Barizi, MA, selaku dosen pembimbing skripsi, yang telah memberikan banyak pengarahan dan pengalaman yang berharga.
5. Abdul Aziz, M.Si, selaku tim penguji skripsi, terimakasih telah memberikan masukan-masukan yang berharga dan bermanfaat untuk penulisan skripsi ini.
6. Seluruh dosen jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah membantu dalam menyelesaikan skripsi ini.
7. Keluarga tercinta yang senantiasa mendo'akan dan memberikan motivasi kepada kami agar mencapai kesuksesan.
8. Teman-teman seperjuangan mahasiswa Matematika Angkatan 2008, terima kasih atas segala pengalaman dan kenangan terindah saat menuntut ilmu bersama.
9. Semua pihak yang tidak mungkin penulis sebutkan satu persatu, terimakasih atas bantuan moral maupun spiritual yang telah diberikan kepada penulis.

Semoga Allah SWT membalas semua amal kebaikan yang telah mereka berikan kepada kami, Amin.

Penulis menyadari sebagai manusia biasa, skripsi ini masih jauh dari kekurangan dan kesempurnaan, penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun dari berbagai pihak agar dalam karya tulis selanjutnya dapat lebih baik dan lebih memberikan manfaat bagi pembaca.

Wassalamu'alaikum Wr.Wb.

Malang, Januari 2012

Penulis,

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SIMBOL

$\rightarrow$	:	menuju
μ	:	nilai tengah (rataan)
$\bar{X}$	:	rata-rata pada pengamatan X
$\bar{Y}$	:	rata-rata pada pengamatan Y
E	:	expectation (nilai harapan)
s^2	:	ragam untuk sampel
σ^2	:	ragam (varian) untuk populasi
T	:	transpose
N	:	normal
$\hat{\beta}^*$	:	estimasi dari parameter $\hat{\beta}$
$\hat{\sigma}^2$	:	estimasi dari parameter σ^2
Z	:	*outier*
$L(x_1, \dots, x_n; \theta)$	:	fungsi *likelihood*
$f_{x_1,\dots,x_n}(X_1, X_2, \dots, X_n; \theta)$	:	fungsi padat peluang
$X_1, X_2, \dots, X_n$	:	peubah acak

ABSTRAK

Ngaini, Nur. 2012. **Estimasi Parameter Model Regresi Linier pada Data yang Mengandung *Outlier* dengan Metode *Maximum Likelihood Estimation***. Skripsi. Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (1) Sri Harini, M.Si. (II) Dr. Ahmad Barizi, MA.

Kata kunci: *outlier*, *maximum likelihood estimation*, regresi linier, estimasi parameter

Secara umum *outlier* dapat diartikan data yang tidak mengikuti pola umum pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan. *Outlier* merupakan salah satu faktor yang dapat mempengaruhi estimasi parameter pada model regresi linier. Untuk mengetahui apakah *outlier* berpengaruh terhadap estimasi parameter pada model regresi linier dilakukan dengan jalan mengestimasi parameter model regresi linier yang mengandung *outlier* dan mengaplikasikan hasil estimasi parameter tersebut pada data yang mengandung *outlier*.

Penelitian ini bertujuan untuk mengestimasi parameter model regresi linier yang mengandung *outlier* dan diharapkan dapat mempermudah para peneliti dalam mengestimasi parameter model regresi linier yang mengandung *outlier*. Metode yang digunakan untuk mengestimasi parameter model regresi linier yang mengandung *outlier* adalah metode *maximum likelihood estimation*. Untuk membuktikan pengaruh *outlier* terhadap suatu estimasi parameter pada model regresi linier dilakukan suatu pengujian terhadap estimasi parameter yang dihasilkan dari metode *maximum likelihood estimation* yaitu dengan cara menentukan sifat-sifat estimasi parameter yang mengandung *outlier* dan menerapkan langsung pada data yang mengandung *outlier*. Setelah itu, menghilangkan data yang terdapat *outlier* dan mengestimasi kembali model tersebut.

Hasil penelitian ini menujukkan bahwa estimasi parameter model regresi linier yang tidak mengandung *outlier* lebih baik daripada estimasi parameter model regresi linier yang mengandung *outlier*. Akan tetapi pada aplikasi datanya, nilai estimasi parameter yang dihasilkan model regresi linier pada data yang mengandung *outlier* lebih kecil daripada nilai estimasi parameter model regresi linier pada data ketika *outlier*nya dihilangkan.

ABSTRACT

Ngaini, Nur. 2012. **Parameter Estimation in Linear Regression Models Containing Data Outlier by Using Maximum Likelihood**. Thesis. Mathematics Programme Faculty of Science and Technology The State of Islamic University Maulana Malik Ibrahim Malang.
Promotor: (1) Sri Harini, M. Si
(II) Dr. Barizi Ahmad, MA

In general, outliers can be interpreted the data that do not follow the general pattern on the model or data out of the model and not in the confidence interval. Outlier is one factor that can affect the estimation of parameters in linear regression models. To determine whether the outlier effect on parameter esrimation in linear regression model was done by estimating parameters of linear regression models containing outlier and apply the result of parameter estimation on data that contain outliers.

This study aims to estimate the parameters of linear regression models containing outlier and is expected to facilitate the reseachers in estimating the parameters of linear regression models that contain outlier. the method used to estimate parameters of linear regression models containing outlier is the maximum likelihood estimation method. To prove the influence of outlier on the estimation of parameters in linear regression models carried out an examination of parameter estimates resulting from maximum likelihood estimation method is by way of determining the properties of parameter estimates wich contain outliers and apply directly on the data contain outliers. After that, remove the data contained outliers and estimate the model again.

The results of this study showed that the estimated parameters of the linear regression model that does not contain outlier is better than linear regression model parameter estimation that contain outlier. However, the application data, the value of the parameter estimates generated linear regression models to data containing outliers is smaller than the value of the parameter estimation of linear regression model to the data when its outlier eliminated.

Keywords: outliers, maximum likelihood estimation,
linear regression, the estimated parameters

الملخّص

نور عيني،٢٠١٢. **تقدير الممثلة للمجتمع ببشكل الترتيب التأخر العمودي على البيانات التي لها الأوتلير (*Outlier*) بطريقة نهاية أكثر الممكنات التقديرية (*Maximum Likelihood Estimation*)**. البحث الجامعي. الشعبة الرياضيات لكلية العلوم والتكنولوجية بجامعة الإسلامية الحكومية مالانج. المشرف : سري هاريني الماجستير و د. أ. بارزي الماجستير

الكلمة الرئيسية : أوتلير، نهاية أكثر الممكنات التقدرية، الترتيب التأخر، تقدير ممثلة للمجتمع

عامة، أنّ أوتلير تعني البيانات التي لاتتبعها أشكال عامة أو من أشكال البيانات سوى أشكالها ولاتوجد حوالي تبادل الصدق. أوتلير من العوامل التي يتأثرها تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي. لمعرفة أم كانت أوتلير أن تتأثر تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي تُستخدَم أن تقدر ممثلة للمجتمع بشكل الترتيب التأخر العمودي التي لها أوتلير و أن تطبّق حاصلها على البيانات التي لها الأوتلير.

هذا البحث تهدف لأن تقدرَ ممثلة للمجتمع بشكل الترتيب التأخر العمودي التي لها أوتلير و تُرجَى أن يسهّل الباحثون في تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي التي لها أوتلير. وأما كانت الباحثة تستخدم الطريقة نهاية أكثر الممكنات التقدرية. ولتحقيق تأثر أوتلير على تقدير الممثلة للمجتمع في بشكل الترتيب التأخر العمودي فاستختدم الإختبار على تقدير الممثلة للمجتمع التي تحصل على تلك الطريقة تعني بأن تعيّن الصفات لها التي لها الأوتلير وتطبّقها في البيانات التي لها أوتلير مباشرة وبعدها أن تمسح البيانات التي لها أوتلير وأن تقدر مرة بعد مرات في شكلها.

حصل هذا البحث على تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي التي ليس لها أوتلير أحسن من تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي التي لها أوتلير. ولكن في تطبيق بياناتها أن قيمة تقدير الممثلة للمجتمع التي تُحصَل على شكل الترتيب التأخر العمودي للبيانات التي لها أوتلير أصغر من قيمة تقدير الممثلة للمجتمع بشكل الترتيب التأخر العمودي للبيانات التي تمسح أوتلير فيها.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Al-Qur’an merupakan sumber ilmu pengetahuan yang tiada tandingannya di muka bumi ini. Al-Qur’an menjelaskan dimensi baru dan aktual yang sangat luas. Al-Qur’an tidak hanya membahas tentang agama saja, melainkan juga membahas tentang sosial, ekonomi bahkan sains pun dibahas di dalamnya.

Dalam bidang matematika, contohnya tentang estimasi dan *outlier* pun juga disinggung dalam Al-Qur’an, yaitu pada surat Ash-Shaaffat ayat 147 yang meyinggung tentang estimasi dan surat Al-Jin ayat 14 yang meyinggung tentang *outlier*, lebih jelasnya adalah sebagai berikut:

وَأَرۡسَلۡنَٰهُ إِلَىٰ مِاْئَةِ أَلۡفٍ أَوۡ يَزِيدُونَ ١٤٧

Artinya: “*Dan Kami utus Dia kepada seratus ribu orang atau lebih*” (Qs. Ash-Shaaffat 37:147).

Surat Ash-Shaaffat ayat 147 tersebut menelaskan bahwa Nabi Yunus diutus kepada umatnya yang jumlahnya 100.000 orang atau lebih. Jika membaca ayat tersebut secara seksama, maka terdapat rasa atau kesan ketidakpastian dalam menentukan jumlah umat Nabi Yunus. Allah tidak menyebutkan umat Nabi Yunus secara jelas dan detail akan tetapi dinyatakan dengan suatu perkiraan. Sehingga dari gambaran di atas dapat diketahui bahwa itulah contoh estimasi dalam Al-Qur’an.

وَأَنَّا مِنَّا ٱلْمُسْلِمُونَ وَمِنَّا ٱلْقَٰسِطُونَ ۖ فَمَنْ أَسْلَمَ فَأُو۟لَٰٓئِكَ تَحَرَّوْا۟ رَشَدًا ﴿١٤﴾

Artinya: "*Dan Sesungguhnya di antara Kami ada orang-orang yang taat dan ada (pula) orang-orang yang menyimpang dari kebenaran. Barangsiapa yang yang taat, Maka mereka itu benar-benar telah memilih jalan yang lurus*" (Qs. Al-Jin 72:14).

Surat Al-Jin ayat 14 di atas, Allah menjelaskan tentang jin bahwa di antara mereka ada yang beriman mentaati Allah, khusyuk dan ikhlas serta beramal saleh karena-Nya. Ada pula di antara mereka yang berpaling dari haluan yang benar. Barang siapa beriman kepada Allah dan mentaati-Nya sesungguhnya dia telah menempuh jalan yang akan menyampaikannya kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkannya dari siksa neraka. Jika ditelaah ayat di atas mejelaskan suatu peyimpangan, layaknya suatu data yang mengalami peyimpangan dari sekumpulan data. Sehingga dari gambaran di atas dapat diketahui bahwa itulah contoh *outlier* dalam Al-Qur'an

Pengamatan *outlier* adalah suatu pengamatan dimana terdapat penyimpangan-penyimpangan dalam sekumpulan data hasil penelitian. Data yang menyimpang dari sekumpulan data yang lain disebut dengan *outlier* Apabila dalam suatu data terdapat *outlier*, dapat menyebabkan nilai error makin besar dan dapat memperkecil atau menurunkan nilai koefisien regresi dan juga nilai korelasi, selain itu dapat menyebabkan data hasil pengamatan tidak menyebar normal.

Banyak masalah praktis yang berhubungan dengan statistika inferensia salah satunya adalah mengenai regresi yang merupakan metode statistika yang paling umum digunakan. Menurut Draper dan Smith, metode regresi yaitu metode yang menghubungkan variabel terikat dengan variabel bebas dengan hasil keluaran utamanya adalah estimasi parameter yang membentuk suatu model tertentu.

Statistik inferensia merupakan proses yang menggunakan sampel statistik untuk mengestimasi hubungan parameter populasi yang tidak diketahui, contohnya acak dan hitung peluang. Dalam hal ini meliputi dua hal yaitu estimasi dan pengujian hipotesis. Pengetahuan tentang hipotesis sangat penting untuk dipelajari begitu juga dengan estimasi yang dipeoleh harus bisa dipertanggung jawabkan.

Pada penelitian ini, akan dilakukan estimasi pada model regresi linier pada suatu data yang mengandung *outlier*. Sehingga dari uraian di atas maka pada penelitian ini akan membahas tentang **Estimasi Parameter Model Regresi Linier pada Data yang Mengandung *Outlier* dengan Metode *Maximum Likelihood Estimation***.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana estimasi parameter model regresi linier yang mengandung *outlier* dengan menggunakan metode *Maximun Likelihood Estimation* ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah mendapatkan estimasi parameter model regresi linier yang mengandung *outlier* dengan menggunakan metode *Maximum Likelihood Estimation*.

1.4 Batasan Masalah

Berdasarkan rumusan masalah dan tujuan penelitian yang telah disebutkan di atas, maka batasan masalah yang diberikan adalah:

1. Estimasi pada parameter β dan σ^2 pada model regresi linier yang mengandung *outlier*.
2. *Outlier* yang terjadi pada variabel bebas (X).
3. Asumsi bahwa model regresi linier mengikuti distribusi normal yaitu $\varepsilon \sim N(0, \sigma^2)$.
4. Model *outlier* pada regresi linier univariat.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

a. Bagi Peneliti

Mengembangkan dan memperdalam ilmu peneliti dalam memberikan pemahaman mengenai estimasi parameter model regresi linier khususnya pada suatu data yang mengandung *outlier* dengan metode *Maximum Likelihood Estimation*.

b. Bagi Pembaca

Sebagai tambahan wawasan, bahan bacaan, referensi dan informasi mengenai estimasi parameter model regresi linier khususnya pada suatu data yang mengandung *outlier* dengan menggunakan metode *Maximun Likelihood Estimation*.

c. Pengembangan Ilmu Matematika

Sebagai sumbangan pemikiran keilmuan matematika khususnya dalam bidang statistik.

1.6 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode kepustakaan yaitu usaha mendalami, mencermati, menelaah dan mengidentifikasi pengetahuan yang ada dalam kepustakaan (sumber bacaan, buku-buku referensi atau hasil penelitian orang lain) sebagai literatur untuk mengumpulkan data-data dan informasi (Hasan, 2002:45).

Menurut Mardalis (1990:28), penelitian kepustakaan bertujuan untuk mengumpulkan data dan informasi dengan bermacam-macam material yang terdapat dalam ruangan perpustakaan, seperti buku, majalah, dokumen catatan dan kisah-kisah sejarah lainnya.

Adapun langkah-langkah dalam penelitian ini adalah:

1. Menentukan model persamaan regresi linier yang mengandung *outlier*.

2. Menentukan fungsi *likelihood* yang diperoleh dari fungsi distribusi peluang, kemudians mengubah bentuk fungsi likelihood menjadi *log-likelihood*.
3. Menentukan estimasi parameter pada model regresi linier yang mengandung *outlier* dengan metode *Maximum Likelihood Estimation* dengan mencari nilai estimasi parameter β dan σ^2.
4. Menentukan sifat-sifat parameternya.
5. Memberikan contoh aplikasi yang ada hubungannya dengan estimasi parameter model regresi linier pada data yang mengandung *outlier* dengan metode *Maximum Likelihood Estimation*.
6. Membuat kesimpulan-kesimpulan yang merupakan jawaban dari permasalahan yang telah dikemukakan pada pembahasan.

1.7 Sistematika Penulisan

Untuk memudahkan melihat dan memahami penelitian ini secara menyeluruh, maka sistematika penulisan skripsi ini dibagi menjadi empat bab yaitu:

BAB I PENDAHULUAN, berisi tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metode penelitian dan sistematika penulisan.

BAB II KAJIAN PUSTAKA, menjelaskan tentang teori-teori yang berkaitan dengan estimasi parameter model regresi linier yang mengandung *outlier* dengan menggunakan metode *Maximun Likelihood Estimation*.

BAB III PEMBAHASAN, Pada bab ini berisi tentang hasil penelitian yang mengkaji estimasi parameter model regresi linier yang mengandung *outlier* dengan menggunakan metode *Maximun Likelihood Estimation* dan menentukan sifat-sifat estimator parameter serta menerapkan aplikasinya.

BAB IV PENUTUP, berisi tentang kesimpulan dan saran-saran yang sesuai dengan hasil penelitian.

BAB II

KAJIAN PUSTAKA

Berikut ini merupakan teori-teori yang berkaitan dengan estimasi parameter model regresi linier pada data yang mengandung *outlier* dengan metode *maximum likelihood estimation*.

2.1 Estimasi Parameter

2.1.1 Pengertian Estimasi Parameter

Estimasi adalah suatu metode untuk mengetahui sekitar nilai-nilai suatu populasi dengan menggunakan nilai-nilai sampel. Nilai-nilai populasi sering disebut dengan perameter populasi, sedangkan nilai-nilai sampel sering disebut dengan statistik sampel. Dalam metode estimasi, parameter populasi yang ingin diestimasi itu adalah berupa nilai rata-rata yang diberi notasi μ dan nilai simpangan baku dengan notasi σ. Dengan menggunakan data sampel maka berusaha untuk mengetahui karakteristik populasi.

Estimasi adalah proses yang menggunakan sampel (statistik) untuk mengestimasi hubungan parameter dengan populasi yang tidak diketahui. Estimasi merupakan suatu pernyataan mengenai parameter populasi yang diketahui berdasarkan dari sampel, dalam hal ini peubah acak yang diambil dari populasi yang bersangkutan. Jadi dengan estimasi ini keadaan parameter populasi dapat diketahui (Hasan, 2002:11).

Menurut Yitnosumarto (1990:211-212) estimasi adalah anggota peubah acak dari statistik yang (anggota peubah diturunkan). Besaran sebagai hasil penerapan estimasi terhadap data dari semua contoh disebut nilai estimasi.

Pada umumnya estimasi parameter menempuh langkah-langkah sebagai berikut:

a. Menetapkan besaran parameter yang akan diestimasi
b. Memilih kerangka estimasi yaitu distribusi sampling yang sejenis dengan besaran parameter yang akan diestimasi
c. Menentukan taraf kepercayaan
d. Proses perhitungan
e. Membuat kesimpulan berdasarkan proses perhitungan

2.1.2 Macam-Macam Estimasi Parameter

Murray dan Larry (1999:166) menyatakan terdapat dua jenis estimasi parameter, yaitu:

1. Estimasi titik

   Estimasi dari sebuah parameter populasi yang dinyatakan oleh bilangan tunggal disebut sebagai estimasi titik dari parameter tersebut. Sebuah nilai yang diperoleh dari sampel dan digunakan sebagai estimasi dari parameter yang nilainya tidak diketahui. Misalkan $X_1, X_2, \dots, X_n$ merupakan sampel acak berukuran n dari X, maka statistik yang berkaitan dengan θ dinamakan estimasi dari θ. Setelah sampel diambil, nilai-nilai yang dihitung dari sampel itu digunakan sebagai taksiran titik bagi θ.

2. Estimasi Interval

Estimasi dari parameter populasi yang dinyatakan dengan dua buah bilangan. diantara posisi parameternya diperkirakan berbeda disebut estimasi interval. Estimasi interval mengindikasikan tingkat kepresisian atau akurasi dari sebuah estimasi sehingga estimasi interval akan dianggap semakin baik jika mendekati estimasi titik.

2.1.3 Sifat-Sifat Estimasi

1. Tak Bias *(Unbiased)*

Satu hal yang menjadi tujuan dalam estimasi adalah estimator harus mendekati nilai sebenarnya dari parameter yang diestimasi tersebut. Misalkan terdapat parameter θ. Jika $\hat{\theta}$ merupakan estimator tak bias (unbiased estimator) dari parameter θ, maka $E(\hat{\theta}) = \theta$ (Yitnosumarto, 1990:212).

Yusuf Wibisono (2005:362) dalam bukunya menyatakan bahwa estimator tak bias bagi parameter θ, jika

$$E(\hat{\theta}) = \theta$$

dan dikatakan estimator bias bagi parameter θ, jika

$$E(\hat{\theta}) \neq \theta$$

Namun estimator bias dapat diubah menjadi estimator tak bias jika ruas kanan dikalikan atau ditambahkan dengan konstanta tertentu.

2. Efisien

Jika distribusi sampling dari dua statistik memiliki mean atau ekspektasi yang sama, maka statistik dengan varians yang lebih kecil disebut sebagai estimator efisien dari mean, sementara statistik yang lain disebut estimator tak efisien. Adapun nilai-nilai yang berkorespondensi dengan statistik-statistik ini masing-masing disebut sebagai estimasi efisien dan estimasi tak efisien.

Suatu estimator (misalkan: $\hat{\theta}$) dikatakan efisien bagi parameter (θ) apabila estimator tersebut mempunyai varians yang kecil. Apabila terdapat lebih satu estimator, estimator yang efisien adalah estimator yang mempunyai varians kecil. Dua estimator dapat dibandingkan efisiensi relatif (relative efficiency). Efisien relatif $\hat{\theta}_2$ terhadap $\hat{\theta}_1$ dirumuskan:

$$R(\hat{\theta}_2, \hat{\theta}_1) = \frac{E(\hat{\theta}_1 - \hat{\theta})^2}{E(\hat{\theta}_2 - \hat{\theta})^2}$$

$$= \frac{E\left(\hat{\theta}_1 - E(\hat{\theta}_1)\right)^2}{E(\hat{\theta}_2 - E(\hat{\theta}_2))^2}$$

$$= \frac{var\ \hat{\theta}_1}{var\ \hat{\theta}_2}$$

$R = \frac{\hat{\theta}_1}{\hat{\theta}_2}$, jika R > 1 maka $\hat{\theta}_1 > \hat{\theta}_2$ artinya secara relatif $\hat{\theta}_2$ lebih efisien daripada $\hat{\theta}_1$ dan jika R < 1 maka $\hat{\theta}_1 < \hat{\theta}_2$ artinya secara relatif $\hat{\theta}_1$ lebih efisien dari pada $\hat{\theta}_2$.

3. Konsisten

Suatu estimator dikatakan konsisten, jika memenuhi syarat sebagai berikut:

1) Jika ukuran sampel semakin bertambah maka estimator akan mendekati parameternya. Jika besar sampel menjadi tak terhingga maka estimator konsisten harus dapat memberi suatu estimator titik yang sempurna terhadap parameternya. Jadi, $(\hat{\theta})$ merupakan estimator konsisten, jika dan hanya jika:

   $E\left(\hat{\theta} - E(\theta)\right)^2 \to 0$ jika $n \to \infty$

2) Jika ukuran sampel bertambah besar maka distribusi sampling penduga akan mengecil menjadi suatu garis tegak lurus di aats parameter yang sama dengan probabilitas sama dengan 1 (Hasan, 2002:113-115).

Gujarati (2007:98) menerangkan estimator parameter $\hat{\theta}$ dikatakan konsisten bila nilai-nilainya mendekati nilai parameter yang sebenarnya meskipun ukuran sampelnya semakin besar. Suatu statistik $\hat{\theta}$ disebut estimator yang konsisten untuk parameter $\hat{\theta}$ jika dan hanya jika $\hat{\theta}$ konvergen dalam probabilitas ke parameter $\hat{\theta}$

atau

$$p\, lim\, \hat{\theta} = \theta$$

Jika $\hat{\theta}_n$ adalah penaksir untuk θ yang didasarkan pada sampel acak berukuran n, maka $\hat{\theta}_n$ dikatakan konsisten bagi parameter θ, jika $lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon = 1)$

Penentuan estimator konsisten ini dapat dilakukan dengan menggunakan ketidaksamaan Chebyshev's, $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < k\sigma \geq 1 - \frac{1}{k^2})$.

2.2 Model Analisis Regresi

Istilah regresi pertama kali diperkenalkan oleh Francis Galton dalam artikelnya "*Family Likeness in Stature*" pada tahun 1886. Studinya ini menghasilkan apa yang dikenal dengan hukum regresi universal tentang tingginya anggota suatu masyarakat. Hukum tersebut menyatakan bahwa distribusi tinggi suatu masyarakat tidak mengalami perubahan yang besar sekali antar generasi. Hal ini dijelaskan Galton pada fakta yang memperlihatkan adanya kecenderungan mundurnya (*regress*) tinggi rata-rata anak dari orang tua dengan tinggi tertentu menuju tinggi rata-rata seluruh anggota masyarakat. Ini berarti terjadi penyusutan kearah keadaan sedang. Tetapi sekarang istilah regresi telah diberikan makna yang jauh berbeda dari yang dimaksudkan oleh Galton. Secara luas sekarang analisis regresi diartikan sebagai suatu analisis tentang ketergantungan suatu variabel kepada variabel lain dalam rangka membuat estimasi atau prediksi dari rata-rata nilai variabel tergantung dengan diketahuinya nilai variabel bebas (Lains, 2003:19).

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 matematik, maka kita dapat memanfaatkan untuk keperluan-keperluan lain misalnya peramalan. Tujuan utama dari analisis regresi adalah mendapatkan estimasi (ramalan) dari suatu variabel dengan menggunakan variabel lain yang diketahui. Analisis regresi mempunyai dua jenis pilihan yaitu regresi linier dan regresi non linier (Wibisono, 2005:529).

Gujarati (2007:115) menyatakan bahwa analisis regresi menyangkut studi tentang hubungan antara satu variabel terikat atau variabel yang dijelaskan dan satu atau lebih variabel lain yang disebut variabel bebas atau variabel penjelas. Selanjutnya model ini dapat digunakan untuk memprediksi nilai variabel terikat apabila diberikan nilai dari variabel bebas. Oleh karena itu, estimasi model yang didapatkan sebaiknya memenuhi kriteria model yang baik sehingga mampu digunakan sebagai prediksi error yang terkecil.

2.3 Regresi Linier

Menurut Andi Supangat (2008:325) regresi linier merupakan suatu persamaan yang menggambarkan hubungan antara variabel terikat dengan variabel bebas, dimana model berhubungan secara linier dengan variabel terikat. Selanjutnya model ini dapat digunakan untuk memprediksi nilai variabel terikat apabila diberikan nilai dari variabel bebas. Oleh karena itu, estimasi model yang didapatkan sebaiknya memenuhi kriteria model yang baik sehingga mampu digunakan sebagai prediksi error yang terkecil.

Sedangkan menurut Hasan (2000:115) regresi linier adalah di mana variabel-variabelnya (variabel bebas, X dan variabel terikat, Y) berpangkat paling tinggi satu dan saling berhubungan secara linier.

2.3.1 Regresi Linier Sederhana

Regresi linier sederhana adalah regresi linier yang hanya melibatkan dua variabel bebas X dan variabel terikat Y. Model regresi linier sederhana X dan Y ditulis dalam bentuk sebagai berikut:

$$Y = \alpha + \beta X + \varepsilon \tag{2.1}$$

dengan

Y : variabel terikat

X : variabel bebas

α : konstanta

β : koefisien regresi

ε : error

2.3.2 Regresi Linier Berganda

Regresi linier berganda adalah regresi yang variabel terikat Y dihubungkan dengan lebih dari satu variabel bebas X. Bentuk umum model regresi linier berganda adalah:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (2.2)$$

dengan:

Y : variabel terikat

X_1, X_2 : variabel bebas

α : konstanta

β_1, β_2 : koefisien regresi

ε : error

2.4 Model Regresi Linier dalam Pendekatan Matriks

Model regresi linier sederhana dapat digeneralisasikan menjadi lebih dari satu atau k variabel. Persamaan bagi model regresi linier dengan k variabel adalah sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \dots, +\beta_k x_k + \varepsilon \qquad (2.3)$$

Bila pengamatan $y, x_1, x_2, \dots, x_k$ dinyatakan masing-masing dengan $y_i, x_{i1}, x_{i2}, \dots, x_{ik}$ dan errornya ε_i, maka persamaanya adalah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad (2.4)$$

dimana $i = 1,2,3, \dots, n$.

Dinotasikan dalam bentuk matriks menjadi:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{2.5}$$

misalkan:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Secara ringkas persamaan (2.5) dapat ditulis sebagai berikut:

$$Y = X\beta + \varepsilon \tag{2.6}$$

dengan:

Y : vektor peubah terikat ukuran $n\ x\ 1$

X : vektor peubah bebas ukuran $n\ x\ (k+1)$

β : vektor parameter ukuran $(k+1)\ x\ 1$

ε : Vektor galat ukuran $n\ x\ 1$

(Sembiring, 1995:134-135)

2.5 *Outlier*

Outlier adalah pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya. Secara umum *outlier* dapat dibedakan menjadi dua, yaitu *outlier* pada pengamatan dan *outlier* pada model linier.

Berdasarkan banyaknya variabel yang dipertimbangkan *outlier* dapat dibedakan menjadi *outlier* pada pengamatan univariat atau multivariat dan *outlier* pada model linier univariat atau multivariat. *Outlier* pada model linier multivariat dapat dibagi atas tiga kategori, yaitu *outlier* terhadap leverage dan error ataupun keduanya.

Outlier dapat diartikan data yang tidak mengikuti pola umum pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 1995:62).

Sedangkan menurut Draper dan Smith (1992:146) sisaan yang merupakan *outlier* adalah yang nilai mutlaknya jauh lebih besar dari pada sisaan lainnya dan bisa jadi terletak tiga atau empat kali simpangan baku atau lebih jauh lagi dari rata-rata sisannya. *Outlier* merupakan suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya.

Sebagaimana dikemukakan oleh Soemartini (2007) bahwa Ferguson mendefinisikan *outlier* sebagai suatu pengamatan yang meyimpang dari sekumpulan pengamatan yang lain. Barnett mendefinisikan *outlier* adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat.

Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh.

2.6 Metode *Maximum Likelihood*

2.6.1 Fungsi *Likelihood*

Definisi 1:

Fungsi *likelihood* dari n variabel random $X_1, X_2 \dots, X_n$ didefinisikan sebagai fungsi kepadatan bersama dari n variabel random. Fungsi kepadatan bersama $f(x_1, \dots, x_n; \theta)$, yang mempertimbangkan fungsi dari θ. Jika $X_1, X_2 \dots, X_n$ adalah sampel acak dari fungsi kepadatan $f(x;\theta)$, maka fungsi *likelihood* adalah $f(x_1;\theta)f(x_2;\theta) \dots f(x_n;\theta)$ (Mood, Graybill and Boes, 1986:278).

Contoh:

Jika $x_1, x_2, \dots, x_n$ adalah sampel acak dari distribusi $x \sim N(0,1)$. Fungsi *likelihood*nya adalah:

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1;\theta)f(x_2;\theta) \dots f(x_n;\theta), \theta \in \Theta$$

Karena berdistribusi normal, maka fungsi $f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\theta)^2}$

Fungsi *likelihood*nya adalah

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \theta) &= f(x_1;\theta)f(x_2;\theta) \dots f(x_n;\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x_1-\theta)^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x_2-\theta)^2} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x_n-\theta)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\left((x_1-\theta)^2+(x_2-\theta)^2+(x_n-\theta)^2\right)} \end{aligned}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_n-\theta)^2}$$

$$= \left(\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_n-\theta)^2}$$

$$= \left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\right) e^{-\frac{1}{2}\sum_{i=1}^{n}(x_n-\theta)^2}$$

Sehingga fungsi *likelihood*nya dapat ditulis sebagai berikut:

$$L(x_1, x_2, \dots, x_n; \theta) = \left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\right) e^{-\frac{1}{2}\sum_{i=1}^{n}(x_n-\theta)^2}$$

2.6.2 Estimasi *Maximum Likelihood*

Maximum Likelihood adalah metode yang dapat digunakan untuk mengestimasi suatu parameter dalam regresi.

Definisi 2:

Andaikan $X_1, X_2, \dots, X_n$ peubah acak dengan fungsi distribusi $F(x_1, x_2, \dots, x_n|\theta)$ dengan $\theta \in \Theta$ yang tidak diketahui dan fungsi *likelihood*nya adalah

$$L(\theta) = \begin{cases} f(x_1, x_2, \dots, x_n|\theta) \text{jika}, F \text{ mempunyai fungsi padat } f \\ p(x_1, x_2, \dots, x_n|\theta) \text{jika}, F \text{ mempunyai fungsi padat } p \end{cases}$$

setiap $\hat{\theta} = \hat{\theta}_n(X_1, X_2, \dots, X_n) \in \Theta$ sehingga

$$L(\tilde{\theta}) = \sup\{L(\theta) : \theta \in \Theta\}$$

disebut *maximum likelihood estimation*.

(Dudewicz dan Misrhra, 1995:412)

Maximum likelihood dapat diperoleh dengan menentukan turunan dari *L* terhadap parameternya dan menyatakannya samadengan nol. Dalam hal ini, akan lebih mudah untuk terlebih dahulu menghitung logaritma dan kemudian menentukan turunannya. Dengan cara ini kita memperoleh:

$$\frac{1}{f(X_1,\theta)}\frac{\partial f(x_1,\theta)}{\partial \theta} + \cdots + \frac{1}{f(X_n,\theta)}\frac{\partial f(x_n,\theta)}{\partial \theta} = 0$$

Penyelesaian dari persamaan ini, untuk θ dalam bentuk x_k, dikenal sebagai estimator *maksimum likelihood* dari θ

2.7 Distribusi

2.7.1 Distribusi Normal

Distribusi normal merupakan model distribusi peluang yang paling banyak digunakan dalam statistika, teutama berbagai penelitian di bidang ilmu-ilmu biologi, fisika, sosial dll. Selain itu distribusi normal juga merupakan salah satu pendekatan penyelesaian yang cukup baik bagi distribusi-distribusi lain, seperti binomial dan poisson.

Distribusi normal pertama kali diperkenalkan oleh Abraham De Moivre seorang ahli metematika berkebangsaan Perancis yang melarikan diri ke Inggris sekitar tahun 1685. Distribusi Normal mempunyai model kurva berbentuk simetris setangkup, menyerupai genta di sekitar satu nilai yang bertepatan dengan puncak kurva yang menjulur ke kiri dan menjulur ke kanan mendekati sumbu datar sebagai asimtotnya (Wibisono, 2005:290-291).

Distribusi Normal adalah fungsi distribusi peubah acak normal, dengan rataan μ dan variansi σ^2. Bila X menyatakan suatu peubah acak kontinu normal dengan parameter populasi μ dan simpanagn baku σ, maka fungsi yang menentukan kurva galat normal dengan rata-rata dan simpangan bakunya adalah:

$$N(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Selain itu, Walpole dan Myers (1995:165) menjelaskan bahwa distribusi normal baku yaitu distribusi peubah acak normal dengan rataan nol dan variansi dengan lambang $N(0,1)$.

2.7.1 Distribusi Peluang Gabungan

Definisi 3:

Jika X dan Y peubah acak, maka peluang terjadinya secara serentak dari X dan Y dinyatakan sebagai $f(x,y)$ disebut Distribusi Peluang Gabungan untuk setiap pasangan (x,y) (Herrhyanto, 2009:5).

Jika X dan Y adalah dua peubah acak diskrit, maka distribusi peluang terjadinya secara serentak atau bersamaan dinyatakan dengan fungsi $f(x,y)$ dan disebut sebagai distribusi peluang gabungan X dan Y.

Fungsi $f(x,y)$ dikatakan distribusi peluang gabungan peubah acak diskrtit X dan Y bila memenuhi:

1. $f(x,y) \geq = 0$ untuk semua (x,y)
2. $\sum_x \sum_y f(x,y) = 1$
3. $P[(X,Y) \in A] = \sum_A \sum f(x,y)$ untuk setiap daerah A di bidang xy

Jika X dan Y adalah dua peubah acak kontinu, maka distribusi peluang terjadinya secara serentak atau bersamaan dinyatakan dengan fungsi $f(x,y)$ dan disebut sebagai distribusi peluang gabungan X dan Y. Fungsi $f(x,y)$ dikatakan fungsi peluang atau distribusi peluang gabungan peubah acak kontinu X dan Y bila memenuhi:

1. $f(x,y) \geq = 0$ untuk semua (x,y)
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dxdy = 1$
3. $P[(X,Y) \in A] = \int \int f(x,y)\, dxdy$

2.8 Kajian Regresi Linier dalam Menentukan Estimasi Parameter dengan Metode *Maximum Likelihood*

Pada kajian regresi linier sebelumnya telah dijelaskan bahwa model regresi linier ada dua yakni model regresi linier sederhana dan model regresi linier berganda. Dari model tersebut akan dicari estimasi parameternya dengan menggunakan *maximum likelihood estimation*. Di sini peneliti mengambil model regresi linier sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{nk} + \varepsilon_i \quad (2.7)$$

dimana $i = 1,2,3,\dots,n$ dan $k = 0,1,2,\dots,n$.

Bentuk umum model regresi linier tersebut dapat diuraikan menjadi:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 X_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

Apabila ditulis dalam bentuk matriks menjadi:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \qquad (2.8)$$

misalkan:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Secara ringkas persamaan (2.8) dapat ditulis:

$$Y = X\beta + \varepsilon \qquad (2.9)$$

dengan:

Y : vektor peubah dependen yang berukuran $n \; x \; 1$

X : matriks peubah independen berukuran $n \; x \; (k + 1)$

β : vektor koefisien regresi $(k + 1)\,x\,1$

ε : vektor eror ukuran $n\,x\,1$

2.8.1 Menentukan Estimasi Parameter

Dari persamaan (2.9) diketahui bahwa $Y = (y_1, y_2, \dots, y_n)^T$ adalah variabel random, karena diasumsikan berdistribusi normal maka $Y \sim N(X\beta, \sigma^2 I)$ dengan $X = (x_{i1}, x_{i2}, \dots, x_{nk})$ dan $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ dimana $i = 1,2,3, \dots, n$ dan I menyatakan matriks ukuran $n\,x\,n$. Sehingga fungsi distribusi peluang gabungan dari Y adalah

$$f(Y|\beta,\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)} \qquad (2.10)$$

Untuk menentukan penduga parameter menggunakan metode *maximum likelihood estimation*, terlebih dahulu ditentukan fungsi *likelihood* yang diperoleh dari fungsi distribusi peluang gabungan pada persamaan (2.10) di atas sebagai berikut:

$$L(\beta,\sigma^2|Y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}$$

$$= \left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\right) e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)} \qquad (2.11)$$

maka *log likelihood*nya adalah

$$ln\ L(\beta,\sigma^2|Y)$$

$$= ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\ e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}\right)$$

$$= ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\right) ln\ e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}$$

$$= -ln\left((2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}\right) + ln\ e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}$$

$$= -\ ln(2\pi)^{\frac{n}{2}} - ln(\sigma^2)^{\frac{n}{2}} + ln\ e^{-\frac{1}{2\sigma^2}(Y^TY-2\beta^TX^TY+\beta^TX^TX\beta)}$$

$$= -\ ln(2\pi)^{\frac{n}{2}} - ln(\sigma^2)^{\frac{n}{2}} - \frac{1}{2\sigma^2}(Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta)$$

$$= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}Y^TY + \frac{1}{2\sigma^2}2\beta^TX^T - \frac{1}{2\sigma^2}\beta^TX^TX\beta \tag{2.12}$$

2.8.1.1 Estimasi Parameter β

Untuk mengestimasi parameter β yang dinotasikan dengan $\hat{\beta}$ yaitu dengan memaksimumkan paersamaan (2.12) terhadap β artinya mendeferensialkan persamaan (2.12) terhadap β dan disamadenganken nol. Mendeferensialkan persamaan (2.12) terhadap β:

$$\frac{\partial \ln L(\beta,\sigma^2|Y)}{\partial\beta}$$

$$= \frac{\partial\left(\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}Y^TY + \frac{1}{2\sigma^2}2\beta^TX^TY - \frac{1}{2\sigma^2}\beta^TX^TX\beta\right)}{\partial\beta}$$

$$= -0 - 0 - 0 + \frac{1}{\sigma^2}X^TY - \frac{1}{2\sigma^2}X^TX\beta - \frac{1}{2\sigma^2}X^TX\beta -$$

$$\frac{1}{2\sigma^2}(\beta^T X^T X)^T$$
$$= \frac{1}{\sigma^2} X^T Y - \frac{1}{\sigma^2}(X^T X\beta) \tag{2.13}$$

Kemudian persamaan (2.13) disamadengankan nol

$$\frac{\partial \ln L(\beta, \sigma^2 | Y)}{\partial \beta} = 0$$

$$\frac{1}{\sigma^2} X^T Y - \frac{1}{\sigma^2}(X^T X\beta) = 0$$

$$-\frac{1}{\sigma^2}(X^T X\beta) = -\frac{1}{\sigma^2} X^T Y$$

$$\frac{1}{\sigma^2}(X^T X\beta) = \frac{1}{\sigma^2} X^T Y$$

$$(X^T X\beta) = X^T Y$$

$$(X^T X)^{-1}(X^T X)\beta = (X^T X)^{-1} X^T Y$$

$$I\beta = (X^T X)^{-1} X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2.14}$$

Jadi estimasi dari parameter β adalah persamaan (2.14) yaitu:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2.15}$$

2.8.1.2 Estimasi Parameter σ^2

Untuk mengestimasi parameter σ^2 yang dinotasikan dengan $\hat{\sigma}^2$ yaitu dengan memaksimumkam persamaan (2.12) terhadap σ^2 artinya mendeferensialkan persamaan (2.12) terhadap σ^2 dan disamadengankan nol.

Mendeferensialkan persamaan (2.12) terhadap σ^2:

$$\frac{\partial \ln L(\beta,\sigma^2|Y)}{\partial\sigma^2}$$

$$= \frac{\partial\left(-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}Y^TY + \frac{1}{2\sigma^2}2\beta^TX^TY - \frac{1}{2\sigma^2}\beta^TX^TX\beta\right)}{\partial\sigma^2}$$

$$= -0 - \frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}Y^TY + \frac{1}{2(\sigma^2)^2}2\beta^TY - \frac{1}{2(\sigma^2)^2}\beta^TX^TX\beta$$

$$= -\frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}Y^TY + \frac{1}{2(\sigma^2)^2}2\beta^TY - \frac{1}{2(\sigma^2)^2}\beta^TX^TX\beta \qquad (2.16)$$

Kemudian persamaan (2.16) disamadengankan nol, sehingga diperoleh

$$\frac{\partial \ln L(\beta,\sigma^2|Y)}{\partial\sigma^2} = 0$$

$$0 = -\frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}Y^TY + \frac{1}{2(\sigma^2)^2}2\beta^TY - \frac{1}{2(\sigma^2)^2}\beta^TX^TX\beta$$

$$\frac{n}{2\sigma^2} = -\frac{1}{2(\sigma^2)^2}Y^TY + \frac{1}{2(\sigma^2)^2}2\beta^TY - \frac{1}{2(\sigma^2)^2}\beta^TX^TX\beta$$

$$\frac{n\sigma^2}{2(\sigma^2)^2} = -\frac{1}{2(\sigma^2)^2}Y^TY + \frac{1}{2(\sigma^2)^2}2\beta^TY - \frac{1}{2(\sigma^2)^2}\beta^TX^TX\beta$$

$$n\sigma^2 = -Y^TY + 2\beta^TY - \beta^TX^TX\beta$$

$$n\sigma^2 = (Y^TY - 2\beta^TY + \beta^TX^TX\beta)$$

$$\sigma^2 = \frac{1}{n}(Y - X\beta)^T(Y - X\beta) \qquad (2.17)$$

Jadi estimasi dari parameter σ^2 adalah persamaan (2.17) yaitu:

$$\hat{\sigma}^2 = \frac{1}{n}(Y - X\beta)^T(Y - X\beta) \qquad (2.18)$$

2.9 Kajian Al-Qur'an dan Hadits tentang Estimasi dan *Outlier*

2.9.1 Ayat Al-Qur’an tentang Estimasi

Dalam Al-Qur’an, estimasi telah disinggung dalam surat Al-Baqarah ayat 80 yaitu:

وَقَالُوا۟ لَن تَمَسَّنَا ٱلنَّارُ إِلَّآ أَيَّامًا مَّعْدُودَةً ۚ قُلْ أَتَّخَذْتُمْ عِندَ ٱللَّهِ عَهْدًا
فَلَن يُخْلِفَ ٱللَّهُ عَهْدَهُۥٓ ۖ أَمْ تَقُولُونَ عَلَى ٱللَّهِ مَا لَا تَعْلَمُونَ ﴿٨٠﴾

Artinya: “*Dan mereka berkata: "Kami sekali-kali tidak akan disentuh oleh api neraka, kecuali selama beberapa hari saja." Katakanlah: "Sudahkah kamu menerima janji dari Allah sehingga Allah tidak akan memungkiri janji-Nya, ataukah kamu hanya mengatakan terhadap Allah apa yang tidak kamu ketahui?*” (Qs. Al-Baqarah 2:80).

Surat Al-Baqarah ayat 80 di atas, dijelaskan bahwa umat Yahudi berkata bahwa mereka tidak akan disentuh oleh api neraka, kecuali beberapa hari saja, selama kita menyembah anak sapi, yaitu empat puluh hari, sesuai dengan sumpah kita. Dan apabila telah habis empat puluh hari, putuslah siksaan terhadap kita. Pada ayat tersebut terdapat ketidakpastian dalam pernyataan jumlah hitungan hari lama orang Yahudi akan disentuh oleh api neraka.

Sehingga terdapat perbedaan penafsiran antara riwayat yang satu dengan yang lain Ada yang mengungkapkan bahwasannya kata-kata beberapa hari saja dimaknai dengan hitungan dimana perbandingan antara hari di dunia dengan hari Yaumul Akhir adalah satu berbanding 1000, bearti makna beberapa hari dimaknai dengan 7 hari saja. Kemudian ada yang memaknai beberapa hari adalah empat hari.

Pada surat Ash-Shaffaat terdapat ayat yang menyinggung masalah matematika, yaitu tentang estimasi. Surat Ash-Shaffaat adalah Makiyah, yakni turun sebelum Nabi hijrah ke Madinah. Ash-Shaffaat berarti yang berbaris-baris, kalimat yang pertama dari ayat yang pertama, yang disebutkan berbaris-baris itu adalah Malaikat-Malaikat Tuhan di alam malakut, yang tidak tahu berapa jutakah bilangannya, kecuali Allah SWT sendiri. Sedangkan bintang dilangit, yang dapat dilihat mata. Sedangkan pasir dipantai yang dapat ditampung tangan. Sedangkan daun dirimba yang dapat dilihat ketika berpucuk, berdaun dan tanggal dari tampuknya, lagi tidak dapat kita manusia menghitungnya, apakah lagi Malaikat yang ghaib (Amrullah, 1981: 106). Surat Ash-Shaffaat ayat 147 tersebut adalah sebagai berikut:

وَأَرْسَلْنَٰهُ إِلَىٰ مِاْئَةِ أَلْفٍ أَوْ يَزِيدُونَ ﴿١٤٧﴾

Artinya: "*Dan Kami utus Dia kepada seratus ribu orang atau lebih*" (Qs. Ash-Shaaffat 37:147).

Penafsiran surat Ash-Shaaffat di atas meyinggung tentang satuan angka. Surat Ash-Shaaffat adalah surat Makiyah yakni turun sebelum Rasulullah hijrah ke Madinah. Ash-Shaffat berarti berbaris-baris. Dinamai dengan Ash-Shaffat (yang bershaf-shaf) ada hubungannya dengan perkataan Ash-Shaaffat yang terletak pada ayat permulaan surat ini yang mengemukakan bagaimana para malaikat yang berbaris di hadapan Tuhannya yang bersih jiwanya, tidak dapat digoda oleh setan. Hal ini hendaklah menjadi I'tibar bagi manusia dalam menghambakan dirinya kepada Allah, yang tidak tahu berapa banyak jumlahnya, kecuali Allah SWT sendiri.

Pada surat Ash-Shaaffat ayat 147 tersebut dijelaskan bahwa Nabi Yunus diutus kepada umatnya yang jumlahnya 100.000 orang atau lebih. Jika membaca ayat tersebut secara seksama, maka terdapat rasa atau kesan ketidakpastian dalam menentukan jumlah umat Nabi Yunus. Mengapa harus menyatakan 100.000 atau lebih? Mengapa tidak menyatakan dengan jumlah yang sebenarnya? Bukankah Allah SWT mengetahui yang ghaib dan yang nyata? Bukankah Allah SWT Maha Mengetahui segala sesuatu, termasuk jumlah umat Nabi Yunus (Abdusysyakir, 2007:153).

Abdusysyakir (2007:155-156), juga mengatakan dalam bukunya bahwa estimasi adalah keterampilan untuk menentukan sesuatu tanpa melakukan proses perhitungan secara eksak. Dalam matematika terdapat tiga jenis estimasi yaitu estimasi banyak atau jumlah (*numerositas*), estimasi pengukuran dan estimasi komputasional. Sebagaimana dijelaskan dalam uraian berikut ini:

1) Estimasi Banyak atau Jumlah

   Estimasi banyak adalah menentukan banyaknya objek tanpa menghitung secara eksak. Objek di sini maknanya sangat luas. Objek dapat bermakna orang, uang, kelereng, titik, dan mobil. Estimasi pada surat Ash-Shaaffat ayat 147 ini adalah estimasi banyak yaitu banyaknya orang.

2) Estimasi Pengukuran

Estimasi pengukuran adalah menentukan ukuran sesuatu tanpa menghitung secara eksak. Ukuran di sini maknanya sangat luas. Ukuran dapat bermakna waktu, panjang, luas, usia dan volume. Ketika melihat orang berjalan tanpa menanyakan tanggal lahirnya, pembaca dapat menebak atau menaksir usianya. Atau pembaca menaksir waktu yang diperlukan untuk melakukan perjalanan dari Malang ke Jakarta menggunakan sepeda motor. Pembaca juga dapat mengestimasi benda hanya melihat suatu bentuknya.

3) Estimasi Komputasional

Estimasi komputasional adalah menentukan hasil suatu operasi hitung tanpa menghitungnya secara eksak. Ketika dimintai menentukan hasil 97 x 23 dalam waktu sepuluh detik, seorang mungkin akan melihat puluhannya saja, sehingga memperoleh hasil 90 x 20 = 1800 inilah estimasi komputasional. Dengan demikian dapat disimpulkan bahwa seseorang mungkin akan menghitung dengan cara membulatkan kepuluhan terdekat.

Selain dua ayat di atas estimasi juga disinggung daalam surat Al-Jaatsiyah Ayat 24

وَقَالُوا۟ مَا هِىَ إِلَّا حَيَاتُنَا ٱلدُّنْيَا نَمُوتُ وَنَحْيَا وَمَا يُهْلِكُنَآ إِلَّا ٱلدَّهْرُ ۚ

وَمَا لَهُم بِذَٰلِكَ مِنۡ عِلۡمٍۖ إِنۡ هُمۡ إِلَّا يَظُنُّونَ ﴿٢٤﴾

Artinya: "*Dan mereka berkata: "Kehidupan ini tidak lain hanyalah kehidupan di dunia saja, kita mati dan kita hidup dan tidak ada yang akan membinasakan kita selain masa", dan mereka sekali-kali tidak mempunyai pengetahuan tentang itu, mereka tidak lain hanyalah menduga-duga saja*" (Qs. Al-Jaatsiyah 45:24).

Surat Al-Jaatsiyah ayat 24 di atas menjelaskan bahwa orang-orang musyrik yang telah disebutkan sebagian sifat mereka berkata, tidak ada kehidupan lagi sesudah kehidupan yang kita alami. Kita mati, kemudian hiduplah anak-anak sesudah kematian kita. Perkataan seperti itu merupakan pendustaan yang tegas dari mereka terhadap kebangkitan dan akhirat. Ringkasnya mereka berkata, yang ada hanyalah dunia ini saja. Suatu kaum mati, kemudian hiduplah yang lain. Tidak ada kebangkitan dan tidak ada kiamat. Dan tidak ada yang membinasakan kita kecuali berjalannya malam dan siang. Jadi lewatnya malam dan siang itulah yang mempengaruhi kebinasaan orang dana mereka menimbulkan setiap peristiwa kepada masa (Al-Maraghi, 1989:290-291).

Dalam menyatakan bahwa kehidupan ini hanyalah kehidupan dunia saja, dan bahwa yang membinasakan adalah masa, mereka tidaklah mempunyai ilmu yang didasarkan kepada akal maupun maqal (kitab). Jadi ringkasnya mereka adalah meyangka, membuat perkiraan saja tanpa adanya hujjah yang dijadikan pegangan (Al-Maraghi, 1989: 292).

Dari ayat yang telah dijelaskan di atas sangat jelas sekali bahwa yang ada kaitanya dengan estimasi adalah kalimat yang berbunyi "*mereka adalah meyangka, membuat perkiraan saja tanpa adanya hujjah yang* *dijadikan pegangan*". Akan tetapi, lain halnya dalam statistik meskipun mengestimasi (memperkirakan) harus mempunyai pegangan dalam arti mengetahui dan paham ilmu-ilmu yang mempelajari hal tersebut.

2.9.2 Ayat Al-Qur'an tentang *Outlier*

Dalam Al-Qur'an *outlier* disinggung pada surat Al-Jin ayat 14 sebagai berikut:

وَأَنَّا مِنَّا ٱلۡمُسۡلِمُونَ وَمِنَّا ٱلۡقَٰسِطُونَۖ فَمَنۡ أَسۡلَمَ فَأُوْلَٰٓئِكَ تَحَرَّوۡاْ رَشَدٗا ١٤

Artinya: "*Dan Sesungguhnya di antara Kami ada orang-orang yang taat dan ada (pula) orang-orang yang menyimpang dari kebenaran. Barangsiapa yang yang taat, maka mereka itu benar-benar telah memilih jalan yang lurus*" (Qs. Al-Jin 72:14).

Surat Al-Jin ayat 14 di atas, Allah menjelaskan tentang jin bahwa di antara mereka ada yang beriman mentaati Allah, khusyuk dan ikhlas serta beramal saleh karena-Nya. Ada pula di antara mereka yang berpaling dari haluan yang benar. Barangsiapa beriman kepada Allah dan mentaati-Nya sesungguhnya dia telah menempuh jalan yang akan menyampaikannya kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkannya dari siksa neraka.

Menurut Sayyid Qutb dalam tafsirnya Fi Zilalil Qur'an menjelaskan bahwa: Sesungguhnya diantara kami (setelah mendengar Al-Qur'an itu) ada golongan menjadi Muslim dan ada pula golongan yang menyeleweng. Oleh karena itu, siapa menjadi muslim, maka merekalah orang-orang yang memilih jalan hidayat.

Sesungguhnya di antara kami ada orang-orang yang taat dan ada pula orang-orang yang menyimpang dari kebenaran) yakni melewati batas disebabkan kekafiran mereka. (Barang siapa yang taat, maka mereka itu benar-benar telah memilih jalan petunjuk) atau menuju ke jalan hidayah.

Setelah diuraikan di atas dapat diambil kesimpulan bahwa yang menjelaskan *outlier* adalah kalimat “*Dan sesungguhnya di antara kami ada orang-orang yang taat dan ada pula orang-orang yang menyimpang dari kebenaran*” dalam artian *outlier* adalah suatu penyimpangan.

Kata penyimpangan dalam surat di atas pada konsep statistika dapat diartikan sebagai *Outlier*. Sebab suatu *outlier* dikatakan sebagai penyimpangan dilihat dari pengetiannya, yaitu:

1. *Outlier* adalah yang nilai mutlaknya jauh lebih besar dari pada sisaan-sisaan lainnya dan bisa jadi terletak tiga atau empat simpangan baku atau lebih jauh dari rata-rata sisaanya.
2. *Outlier* adalah suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data yang lainnya (Draper dan Smith, 1992:146).
3. *Outlier* adalah data yang tidak mengikuti pola umum model (Sembiring, 1995:62).

Penafsiran ayat ini menjelaskan bahwa para penyimpang yakni mereka yang telah sangat jauh dari kebenaran lagi sangat mantap kekufurannya. Penyimpangan ini mempunyai arti yang sama dengan *outlier* yaitu sama-sama terlaetak sangat jauh diantara data dalam model tersebut.

Sedangkan menurut tafsir ibnu katsir jilid 8 (2007:310) dijelaskan bahwa diantara hamba-hamba Allah yang hidup di alam semesta ini adalah ada yang muslim ada juga yang melakukan penyimpangan. Maksudnya di sini adalah mereka melakukan penyimpangan terhadap kebenaran Allah. Berarti mereka jauh dari kebenran-kebenaran Allah.

Dari ayat-ayat Al-Qur'an (Al-Baqarah, Ash-Shaaffat, Al-Jaatsiyah dan Al-Jin) yang telah disebutkan di atas tadi, dapat diketahui bahwa Allah SWT adalah zat yang ahli segalanya melebihi ahli-ahli dan pakar-pakar ilmu lainya. Jadi, jika di bumi Allah ini terdapat ilmu matematika, maka Allah adalah ahlinya yang paling mengetahui. Dialah Allah zat ahli matematika (matematisi) yang serba Maha. Kalau di bumi Allah ada ilmu fisika maka Allah yang paling mengetahui tentang fisika. Tidak ada yang tidak diketahui Allah SWT. Tidak ada yang tersembunyi bagi Allah SWT sesuatupun yang terjadi di bumi bahkan di langit (Abdusysyakir, 2007:91-92).

2.9.3 Hadits tentang Estimasi

Metode estimasi juga disebutkan dalam suatu hadits pada bab jual-beli sebagai berikut:

Diriwayatkan Malik, dari Nafi’, dari Abdullah bin Umar

عَنْ مَا لِكٍ عَنْ نَا فِعٍ عَن عَبدِ اللهِ بْنِ عُمَرَ رَضِيَ الله عَنهُمَا أَنّ رَسُولُ اللهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ نَهَى عَنِ اْ لمُزَا بَنَةِ, وَاْ لمُزَا بَنَةُ اشْتِرَاءُ الثَّمَرِ بِا لتَّمْرِ كَيْلاً وَبَيْعُ ا لكَرْمِ بِا لزَ بِيْبِ كَيْلاً

“Dari Malik, dari Nafi’, dari Abdullah bin Umar ra bahwa Rasulullah SAW melarang jual beli muzabanah. Muzabanah aalah membeli kurma basah [yang masih berada di atas pohon] dengan kurma kering berdasarkan takaran, dan menjual anggur yang basah dengan anggur kering berdasarkan takaran” (Shahih Bukhari, 2185).

Makna asal “Muzabanah” adalah menolak dengan keras. Atas dasar ini maka peperangan dinamakan “Zabuun”, yakni karena dahsyatnya usaha untuk saling mempertahankan diri dari kedua belah pihak. Adapun penyebab salah satu jenis jual-beli dinamakan “Muzabanah” adalah karena masing-masing dari kedua belah pihak menolak hak yang lain. Atau, karena salah satu dari keduanya apabila tidak puas dan merasa ditipu kemudian ingin membatalkan jual beli, maka pihak yang lain menolak keinginan itu dan tidak mau membatalkannya (Al Asqalani, 2007:308).

“*Yaitu menjual kurama kering dengan kurma basah*”, maksudnya, kurma yang belum matang. Ini adalah pengertian asal jual beli muzabanah. Imam Syafi’I memasukkan semua jual-beli (barter) barang dengan barang yang telah diketahui kadarnya, atau dengan barang yang telah diketahui kadarnya tetapi termasuk barang yang berlaku riba di dalamnya. Dia berkata, “Adapun perkataan orang yang mengatakan ‘Aku menjamin untukmu bahwa buah kurmamu ini akan menghasilkan 20 sha’ (misalnya).

Apabila lebih, maka itu untukku. Tetapi jika kurang dari itu, maka itu menjadi resiko bagiku', maka ini termasuk undian (judi) dan bukan muzabanah (Al Asqalani, 2007:308).

Ibnu Hajar *mengatakan* bahwa pada bab "Menjual Anggur Kering dengan Anggur Kering" melalui jalur Ayyub dari Nafi', dari Ibnu Umar diriwayatkan, "*Muzabanah adalah seseorang menjual buah berdasarkan takaran [Dan dia mengatakan] apabila lebih, maka ia adalah untukku dan apabila kurang, maka ia menjadi resiko bagiku*". Riwayat ini menjelaskan bahwa bentuk perjudian seperti ini dikategorikan pula sebagai jual-beli *muzabanah*. Keberadaannya sebagai salah satu bentuk perjudian tidak menghalangi untuk dimasukkan sebagai jual-beli muzabanah (Al Asqalani, 2007:309).

Imam Malik berkata, "Muzabanah" adalah segala sesuatu yamg tidak diketahui ukurannya, baik berdasarkan takaran, timbangan maupun jumlahnya apabila dijual dengan sesuatu yang diketahui ukurannya, baik berdasarkan takaran ataupun yang lainnya. Dalam hal ini, sama saja apakah ia termasuk barang yang berlaku padanya hukum riba atau barang yang lainnya jika dilakukan secara tunai, sebab dilarangnya jual-beli seperti ini adalah karena telah dimasuki unsur judi dan penipuan (Al Asqalani, 2007: 309).

وَالْمُزَا بَنَةُ اشْتِرَاءُ الثَّمَرِ بِا لتَّمْرِفِى رُءُوْسِ النَّخْلِ "*Muzabanah adalah menjual kurma basah dengan kurna kering yang masih berada di atas pohon*". Ibnu Al-Mahdi menambahkan كَيْلاً "*berdasarkan takaran*". Hal ini berdasarkan hadits Ibnu Umar. Penyebutan kata “takaran” bukan untuk membatasi, tetapi merupakan bentuk jual-beli yang ada pada saat itu. Oleh sebab itu ia tidak mengandung makna implisit, karena disebutkan berdasarkan sebab tertentu. Dari hadits ini dapat diketahui bahwa standar ukuran kurma dan anggur kering adalah takaran (Al Asqalani, 2007:314).

Muzabanah yaitu menjual buah berdasarkan takaran, kaitanya dengan estimasi bahwasanya buah tersebut dijual dengan cara di taksir atau dengan memperkirakan. Dalam artian buah kurma yang masih berada di atas pohon dijual dengan cara diestimasi seharga atau sejumlah dengan buah kurma kering dan begitu juga dengan buah anggur basah dijual dengan cara diestimasi seharga atau sejumlah dengan buah anggur kering.

Diriwayatkan Malik, dari Nafi’, dari Ibnu Umar dari Zaid bin Tsabit

عَنْ مَا لِكٍ نَا فِعٍ عَنْ ابْنِ عُمَرَ عَنْ زَيْدِ بْنِ ثَا بِتٍ رَضِيَ اللهُ عَنْهُمْ أَنَّ رَسُولُ اللهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ أَ رْخَصَ لِصَا حِبِ الْعَرِيَّةِ أَن يَبِيعَهَا بِخَرْ صِهَا

“*Dari Malik, dari Nafi’, dari Ibnu Umar, dari Zaid bin Tsabit sesungguhnya Rasulullah SAW memberi keringanan kepada mereka yang mempunyai ariyah untuk menjualnya dengan kira-kira (ditaksir*)” (Shahih Bukhari, 2188).

Para ulama salaf berbeda pendapat. Apakah anggur atau selainnya masuk kategori kurma dalam hal *ariyah* ?. Sebagian mengatakan tidak diikutkan dari madzhab Azh-Zhahiri. Sebagian pendapat lagi mengatakan diikutkan, ini adalah pendapat yang masyhur dalam madzhab Syafi’i. Ada yang berpendapat bahwa semua buah-buahan dan semua yang dapat disimpan lama dapat diikutkan di dalamnya, ini adalah pendapat madzhab Maliki (Al Asqalani, 2007:312).

رَخَّصَ بَعْدَ ذَلِكَ فِى بَيْعِ الْعَرِيَّةِ "*Beliau memberi keringanan setelah itu pada jual-beli ariyah*", yakni setelah adanya larangan menjual (barter) kurma kering dengan kurma basah. Riwayat ini merupakan dalil paling tegas untuk menolak pemahaman sebagian ulama, yang memahami hadits ini berlaku secara umum pada segala bentuk jual-beli (barter) kurma basah dengan kurma kering termasuk jual beli *ariyah* (Al Asqalani, 2007:312).

بِالرُّطَبِ أَوْ بِا لتَّمرِ "*dengan kurma basah atau kurma kering*". Demikian disebutkan dalam riwayat Imam Bukhari dan Muslim dengan kata "atau". Hal ini memiliki kemungkinan makna *takhyir* atau (pilihan) atau *syak* (keraguan). بِالرُّطَبِ أَوْ بِا لتَّمرِ وَلَمْ يُرخِّصفِى غَيْرِهِ "*dengan kurma basah dan kurma kering, dan beliau tidak memberi keringanan pada selain itu*". An-Nasa'i dan Ath-Thabrani menyebutkan dengan menggunakan kata "dan". Hal ini memperkuat kemungkinan bahwa kata "atau" pada riwayat terdahulu bermakna *takhyir* bukan *syak* (Al Asqalani, 2007:312).

أَن يَبِيعَهَا بِخَرْ صِهَا "*untuk dijual sesuai taksirannya*". Ath-Thabrani menambahkan dengan كَيْلاً "*berdasarkan takaran*". Imam Muslim meriwayatkan dari Yahya bin Yahya dari Malik, "*berdasarkan taksirannya setelah menjadi kurma kering*". Dalam riwayat Imam Muslim "*memberi keringanan dalam jual-beli ariyah, diambil oleh penghuni rumah berdasarkan taksirannya setelah menjadi kurma kering yang mana mereka memakannya dalam keadaan masih basah*". Yahya berkata, "*Ariyah* adalah seseorang membeli kurma kering dan menukarnya dengan kurma basah miliknya dengan memperkirakan atau menaksir berapa banyak jumlahnya setelah kering" (Al Asqalani, 2007:314).

Ariyah adalah membeli kurma kering dan menukarnya dengan kurma basah dengan memperkirakan atau menaksir berapa banyak jumlahnya setelah kering. Kaitanya dengan estimasi yaitu adanya kata "memperkirakan atau menaksir" artinya buah kurma kering tersebut ditukar dengan cara diestimasi berapa banyak jumlahnya atau harganya atau ukurannya dengan buah kurma basah setelah kurma basah tersebut menjadi buah kurma kering.

BAB III

PEMBAHASAN

Pada bab ini akan dibahas tentang model regresi linier pada data yang mengandung *outlier* dengan metode *maximum likelihood estimation* untuk mengestimasi parameter pada model tersebut.

3.1 Menentukan Model Regresi Linier yang Mengandung *Outlier*

Diasumsikan model regresi linier yang dipakai adalah model regresi linier yang mengandung *outlier* adalah

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1(x_{i1} + z_{i1}) + \beta_2(x_{i2} + z_{i2}) + \cdots + \beta_k(x_{nk} + z_{nk}) + \varepsilon_i \\
&= \beta_0 + \beta_1 x_{i1} + \beta_1 z_{i1} + \beta_2 x_{i2} + \beta_2 z_{i2} + \cdots + \beta_k x_{nk} + \beta_k z_{nk} + \varepsilon_i \qquad (3.1)
\end{aligned}
$$

dimana $i = 1,2,3,\dots,n$ dan $k = 0,1,2,\dots,n$.

Dalam bentuk matriks persamaan di atas menjadi:

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} 1 & z_{11} & \dots & z_{1k} \\ 1 & z_{21} & \dots & z_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & \dots & z_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \qquad (3.2)
$$

Misalkan:

$$
Y^* = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}
$$

$$
\beta^* = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad Z^* = \begin{pmatrix} 1 & z_{11} & \dots & z_{1k} \\ 1 & z_{21} & \dots & z_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & \dots & z_{nk} \end{pmatrix}
$$

$$\varepsilon^* = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Persamaan (3.2) dapat dinyatakan sebagai berikut:

$$Y^* = X\beta^* + Z^*\beta^* + \varepsilon^* \tag{3.3}$$

dengan:

Y^* : vektor peubah terikat ukuran $n \ x \ 1$

X : matriks peubah bebas ukuran $n \ x \ (k+1)$

β^* : vektor parameter ukuran $(k+1) \ x \ 1$ yang tak diketahui

Z^* : vektor yang mengandung *outlier* ukuran $n \ x \ (k+1)$

ε^* : vektor galat ukuran $n \ x \ 1$

dimana Z^* merupakan persamaan yang mengandung *outlier*. Sehingga persamaan (3.2) dapat ditulis secara ringkas sebagai berikut:

$$Y^*_{n\,x\,1} = X_{n\,x\,(k+1)}\ \beta^*_{(k+1)\,x\,1}\ Z^*_{n\,x\,(k+1)}\ \beta^*_{(k+1)\,x\,1} + \varepsilon^*_{n\,x\,1} \tag{3.4}$$

3.2 Menentukan Estimasi Parameter Model Regresi Linier yang Mengandung *Outlier*

Dari persamaan (3.4) diketahui bahwa $Y^* = (y_1, y_2, \dots, y_3)^T$ adalah variabel random dan Y^* diasumsikan berdistribusi normal maka $Y^* \sim N(X\beta^* + Z^*\beta^*, \sigma^2 I)$. Untuk memodelkan menjadi fungsi *log Likelihood,* maka terlebih dahulu ditentukan fungsi *likelihood*nya.

Fungsi distribusi peluang gabungan Y^* adalah

$f(Y^*|\beta^*, \sigma^2)$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\left((Y^* - (X\beta^* + Z^*\beta^*))^T (Y^* - (X\beta^* + Z^*\beta^*))\right)}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}((Y^* - X\beta^* - Z^*\beta^*)^T (Y^* - X\beta^* - Z^*\beta^*))}$$

(3.5)

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - \beta^* X - \beta^* Z^*)\right)}$$

sehingga fungsi *likelihood* dari distribusi peluang di atas adalah

$L(\beta^*, \sigma^2|Y^*)$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - X\beta^* - Z^*\beta^*)\right)}$$

$$= \left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\right) e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - X\beta^* - Z^*\beta^*)\right)} \quad (3.6)$$

dan *log likelihoodnya* adalah

$ln\, L(\beta^*, \sigma^2|Y^*)$

$$= ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - X\beta^* - Z^*\beta^*)\right)}\right)$$

$$= ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\right) + ln\, e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - X\beta^* - Z^*\beta^*)\right)}$$

$$= -ln\left((2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}\right) + ln\, e^{-\frac{1}{2\sigma^2}\left(\left(Y^{*T} - \beta^{*T}X^T - \beta^{*T}Z^{*T}\right)(Y^* - X\beta^* - Z^*\beta^*)\right)}$$

$$= -\,ln(2\pi)^{\frac{n}{2}} - ln(\sigma^2)^{\frac{n}{2}} +$$

$$ln\, e^{-\frac{1}{2\sigma^2}\left(Y^{*T}Y^* - 2\beta^{*T}X^TY^* - 2\beta^{*T}Z^{*T}Y^* + \beta^{*T}X^TX\beta^* + 2\beta^{*T}Z^{*T}X\beta^* + \beta^{*T}Z^{*T}Z^*\beta^*\right)}$$

$$= -\,ln(2\pi)^{\frac{n}{2}} - ln(\sigma^2)^{\frac{n}{2}} - \frac{1}{2\sigma^2}(Y^{*T}Y^* - 2\beta^{*T}X^TY^* - 2\beta^{*T}Z^{*T}Y^* +$$

$$\beta^{*T}X^TX\beta^* + 2\beta^{*T}Z^{*T}X\beta^* + \beta^{*T}Z^{*T}Z^*\beta^*)$$

$$= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}Y^{*T}Y^* + \frac{1}{2\sigma^2}2\beta^{*T}X^TY^* + \frac{1}{2\sigma^2}2\beta^{*T}Z^{*T}Y^* - \frac{1}{2\sigma^2}\beta^{*T}X^TX\beta^* - \frac{1}{2\sigma^2}2\beta^{*T}Z^{*T}X\beta^* - \frac{1}{2\sigma^2}\beta^{*T}Z^{*T}Z^*\beta^*$$

$$= -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}Y^{*T}Y^* + \frac{1}{\sigma^2}\beta^{*T}X^TY^* + \frac{1}{\sigma^2}\beta^{*T}Z^{*T}Y^* - \frac{1}{2\sigma^2}\beta^{*T}X^TX\beta^* - \frac{1}{\sigma^2}\beta^{*T}Z^{*T}X\beta^* - \frac{1}{2\sigma^2}\beta^{*T}Z^{*T}Z^*\beta^* \qquad (3.7)$$

3.2.1 Estimasi Parameter β^*

Untuk mengestimasi parameter β^* yang dinotasikan dengan $\hat{\beta}^*$ yaitu dengan memaksimumkan persamaan (3.7) terhadap β^*, artinya mendeferensialkan persamaan (3.7) terhadap β^* dan disamadengankan nol.

Dideferensialkan terhadap β^*:

$$\frac{\partial \ln L(\beta^*, \sigma^2|Y^*)}{\partial \beta^*}$$

$$= -0 - 0 - 0 + \frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* - \frac{1}{2\sigma^2}\left(X^TX\beta^* + (\beta^{*T}X^TX)^T\right) - \frac{1}{\sigma^2}\left(Z^{*T}X\beta^* + (\beta^{*T}X^TZ^*)^T\right) - \frac{1}{2\sigma^2}\left(Z^{*T}Z^*\beta^* + (\beta^{*T}Z^{*T}Z^*)^T\right)$$

$$= \frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* - \frac{1}{2\sigma^2}\left(X^TX\beta^* + X^TX\beta^*\right) - \frac{1}{\sigma^2}\left(Z^{*T}X\beta^* + Z^{*T}X\beta^*\right) - \frac{1}{2\sigma^2}\left(Z^{*T}Z^*\beta^* + Z^{*T}Z^*\beta^*\right)$$

$$= \frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* - \frac{1}{2\sigma^2}2X^TX\beta^* - \frac{1}{\sigma^2}2Z^{*T}X\beta^* - \frac{1}{2\sigma^2}2Z^{*T}Z^*\beta^*$$

$$= \frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* - \frac{1}{\sigma^2}X^TX\beta^* - \frac{1}{\sigma^2}2Z^{*T}X\beta^* - \frac{1}{\sigma^2}Z^{*T}Z^*\beta^* \qquad (3.8)$$

Kemudian persamaan di atas disamadengankan nol

$$\frac{\partial \ln L(\beta^*, \sigma^2|Y^*)}{\partial \beta^*} = 0$$

$$\frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* - \frac{1}{\sigma^2}X^TX\beta^* - \frac{1}{\sigma^2}2Z^{*T}X\beta^* - \frac{1}{\sigma^2}Z^{*T}Z^*\beta^* = 0$$

$$\frac{1}{\sigma^2}X^TY^* + \frac{1}{\sigma^2}Z^{*T}Y^* = \frac{1}{\sigma^2}X^TX\beta^* + \frac{1}{\sigma^2}2Z^{*T}X\beta^* + \frac{1}{\sigma^2}Z^{*T}Z^*\beta^*$$

$$X^TY^* + Z^{*T}Y^* = X^TX\beta^* + 2Z^{*T}X\beta^* + Z^{*T}Z^*\beta^*$$

$$(X^T + Z^{*T})Y^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^*$$

$$(X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^* = (X^T + Z^{*T})Y^*$$

$$(X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^*$$
$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})Y^*$$

$$I\beta^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})Y^*$$

$$\beta^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})Y^* \quad (3.9)$$

Jadi estimasi dari parameter β^* adalah persamaan (3.9) yaitu:

$$\hat{\beta}^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})Y^* \quad (3.10)$$

Estimasi parameter pada persamaan (3.10) dikatakan sebagai estimasi parameter $\hat{\beta}^*$ yang mengandung *outlier*.

3.2.2 Estimasi Parameter σ^2

Untuk mengestimasi parameter σ^2 yang dinotasikan dengan $\hat{\sigma}^2$ yaitu dengan memaksimumkan persamaan (3.7) terhadap σ^2 artinya mendeferensialkan persamaan (3.7) terhadap σ^2 dan disamadengankan nol.

Dideferensialkan terhadap σ^2:

$$ln\, L(\beta^*, \sigma^2|Y^*)$$

$$= -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} Y^{*T} Y^* + \frac{1}{\sigma^2} \beta^{*T} X^T Y^* + \frac{1}{\sigma^2} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{2\sigma^2} \beta^{*T} X^T X \beta^* - \frac{1}{\sigma^2} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{2\sigma^2} \beta^{*T} Z^{*T} Z^* \beta^*$$

$$\frac{\partial \ln L(\beta^*, \sigma^2|Y^*)}{\partial \sigma^2}$$

$$= -0 - \frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2} Y^{*T} Y^* + \frac{1}{(\sigma^2)^2} \beta^{*T} X^T Y^* + \frac{1}{(\sigma^2)^2} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{2(\sigma^2)^2} \beta^{*T} X^T X \beta^* - \frac{1}{(\sigma^2)^2} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{2(\sigma^2)^2} \beta^{*T} Z^{*T} Z^* \beta^*$$

$$= -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} Y^{*T} Y^* + \frac{1}{\sigma^4} \beta^{*T} X^T Y^* + \frac{1}{\sigma^4} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{2\sigma^4} \beta^{*T} X^T X \beta^* - \frac{1}{\sigma^4} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{2\sigma^4} \beta^{*T} Z^{*T} Z^* \beta^* \qquad (3.11)$$

Kemudian persamaan di atas disamadengankan nol

$$\frac{\partial \ln L(\beta^*, \sigma^2|Y^*)}{\partial \sigma^2} = 0$$

$$0 = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} Y^{*T} Y^* + \frac{1}{\sigma^4} \beta^{*T} X^T Y^* + \frac{1}{\sigma^4} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{2\sigma^4} \beta^{*T} X^T X \beta^* - \frac{1}{\sigma^4} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{2\sigma^4} \beta^{*T} Z^{*T} Z^* \beta^*$$

$$= -\frac{n}{\sigma^2} - \frac{1}{\sigma^4} Y^{*T} Y^* + \frac{2}{\sigma^4} \beta^{*T} X^T Y^* + \frac{2}{\sigma^4} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{\sigma^4} \beta^{*T} X^T X \beta^* - \frac{2}{\sigma^4} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{\sigma^4} \beta^{*T} Z^{*T} Z^* \beta^*$$

$$\frac{n}{\sigma^2} = \qquad -\frac{1}{\sigma^4} Y^{*T} Y^* + \frac{2}{\sigma^4} \beta^{*T} X^T Y^* + \frac{2}{\sigma^4} \beta^{*T} Z^{*T} Y^* -$$

$$\frac{1}{\sigma^4} \beta^{*T} X^T X \beta^* - \frac{2}{\sigma^4} \beta^{*T} Z^{*T} X \beta^* - \frac{1}{\sigma^4} \beta^{*T} Z^{*T} Z^* \beta^*$$

$$\frac{n\sigma^2}{\sigma^4} = \frac{1}{\sigma^4}\left(-Y^{*T}Y^* + 2\beta^{*T}X^TY^* + 2\beta^{*T}Z^{*T}Y^* - \beta^{*T}X^TX\beta^* - 2\beta^{*T}Z^{*T}X\beta^* - \beta^{*T}Z^{*T}Z^*\beta^*\right)$$

$$\frac{1}{\sigma^4}(n\sigma^2) = \frac{1}{\sigma^4}\left(-Y^{*T}Y^* + 2\beta^{*T}X^TY^* + 2\beta^{*T}Z^{*T}Y^* - \beta^{*T}X^TX\beta^* - 2\beta^{*T}Z^{*T}X\beta^* - \beta^{*T}Z^{*T}Z^*\beta^*\right)$$

$$n\sigma^2 = -Y^{*T}Y^* + 2\beta^{*T}X^TY^* + 2\beta^{*T}Z^{*T}Y^* - \beta^{*T}X^TX\beta^* - 2\beta^{*T}Z^{*T}X\beta^* - \beta^{*T}Z^{*T}Z^*\beta^*$$

$$= Y^{*T}Y^* - 2\beta^{*T}X^TY^* - 2\beta^{*T}Z^{*T}Y^* + \beta^{*T}X^TX\beta^* + 2\beta^{*T}Z^{*T}X\beta^* + \beta^{*T}Z^{*T}Z^*\beta^*$$

$$= (Y^* - X\beta^* - Z^*\beta^*)^T(Y^* - X\beta^* - Z^*\beta^*)$$

$$= \left(Y^* - (X\beta^* + Z^*\beta^*)\right)^T\left(Y^* - (X\beta^* + Z^*\beta^*)\right)$$

$$\sigma^2 = \frac{1}{n}\left(Y^* - (X\beta^* + Z^*\beta^*)\right)^T\left(Y^* - (X\beta^* + Z^*\beta^*)\right) \tag{3.12}$$

Jadi estimasi dari parameter σ^2 adalah persamaan (3.12) yaitu:

$$\hat{\sigma}^2 = \frac{1}{n}\left(Y^* - (X\beta^* + Z^*\beta^*)\right)^T\left(Y^* - (X\beta^* + Z^*\beta^*)\right) \tag{3.13}$$

Estimasi parameter pada persamaan (3.13) dikatakan sebagai estimasi parameter σ^2 yang mengandung *outlier*.

3.3 Menentukan Sifat-Sifat Estimasi Parameter Regresi Linier Yang Mengandung *Outlier*

Persamaan yang mengandung *outlier* $Y^* = X\beta^* + Z^*\beta^* + \varepsilon^*$ diasumsikan $Y^* \sim N(\beta^*X + \beta^*Z^*, \sigma^2 I)$, sehingga $E(Y^*) = X\beta^* + Z^*\beta^*$

dan $var(Y^*) = \sigma^2 I$ dari persamaan tersebut juga diasumsikan bahwa ε variabel bebas berdistribusi normal $\varepsilon^* \sim N(0, \sigma^2)$ karena

$$
\begin{aligned}
E(\varepsilon^*) &= E\big(Y^* - (X\beta^* + Z^*\beta^*)\big) \\
&= E(Y^*) - E(X\beta^* + Z^*\beta^*) \\
&= (X\beta^* + Z^*\beta^*) - (X\beta^* + Z^*\beta^*) \\
&= (X\beta^* + Z^*\beta^*) - (X\beta^* + Z^*\beta^*) \\
&= 0 \qquad (3.14)
\end{aligned}
$$

Karena X dan Z merupakan suatu tetapan dan tidak mempunyai distribusi maka dapat ditentukan sifat-sifat estimasi parameternya, sebagai berikut:

3.3.1 Tak Bias (*Unbias*)

Estimator β^* dikatakan estimator tak bias karena $E(\hat{\beta}^*) = \beta^*$.

Bukti:

$$
\begin{aligned}
E(\hat{\beta}^*) &= E\big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^T + Z^{*T})Y^* \\
&= E\big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^T + Z^{*T})E(Y^*) \\
&= E\big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^T + Z^{*T})E(Y^*) \\
&= \big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^T + Z^{*T})\, X\beta^* + Z^*\beta^* \\
&= \big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}\big(X^T + Z^{*T}\big)(X + Z^*)\beta^* \\
&= \big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^TX + X^TZ^* + Z^{*T}X + Z^{*T}Z^*)\beta^* \\
&= \big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^TX + Z^{*T}X + Z^{*T}X + Z^{*T}Z^*)\beta^* \\
&= \big(X^TX + 2Z^{*T}X + Z^{*T}Z^*\big)^{-1}(X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^*
\end{aligned}
$$

$$= I\beta^*$$

$$= \beta^* \tag{3.15}$$

Dari persamaan (3.15) diperoleh $E(\hat{\beta}^*) = \beta^*$ maka $\hat{\beta}^*$ yang mengandung *outlier* merupakan estimator tak bias.

3.3.2. Efisien

Suatu estimator dikatakan efisien apabila estimator tersebut mempunyai variansi yang kecil.

Perhatikan bahwa:

$$Var(\hat{\beta}^*) = E\left[\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)^T\right]$$

$$= E\left[(\hat{\beta}^* - \beta^*)(\hat{\beta}^* - \beta^*)^T\right]$$

karena

$$\hat{\beta}^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})Y^*$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})(X\beta^* + Z^*\beta^* + \varepsilon^*)$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})((X + Z^*)\beta^* + \varepsilon^*)$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}\left((X^T + Z^{*T})(X + Z^*)\beta^* + (X^T + Z^{*T})\varepsilon^*\right)$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}\left((X^TX + X^TZ^* + Z^{*T}X + Z^{*T}Z^*)\beta^* + (X^T + Z^{*T})\varepsilon^*\right)$$

$$(X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}\left((X^TX + Z^{*T}X + Z^{*T}X + Z^{*T}Z^*)\beta^* + (X^T + Z^{*T})\varepsilon^*\right)$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}\left((X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^* + (X^T + Z^{*T})\varepsilon^*\right)$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^TX + 2Z^{*T}X + Z^{*T}Z^*)\beta^*$$

$$+ (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*$$

$$= I\beta^* + (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*$$

$$= \beta^* + (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*$$

$$\hat{\beta}^* - \beta^* = (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*$$

maka

$$Var(\hat{\beta}^*) = E\left[(\hat{\beta}^* - \beta^*)(\hat{\beta}^* - \beta^*)^T\right]$$

$$= E\left[\left((X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*\right)\right.$$

$$\left.\left((X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*\right)^T\right]$$

$$= E\left[(X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\varepsilon^*\varepsilon^{*T}(X + Z^*)\right.$$

$$\left.(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\right]$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})E[\varepsilon^*\varepsilon^{*T}]$$

$$(X + Z^*)(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})\sigma^2 I$$

$$(X + Z^*)(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^T + Z^{*T})(X + Z^*)$$

$$(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^TX + X^TZ^* + Z^{*T}X + Z^{*T}Z^*)$$

$$(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^TX + Z^{*T}X + Z^{*T}X + Z^{*T}Z^*)$$

$$(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$$

$$= (X^TX + 2Z^{*T}X + Z^{*T}Z^*)^{-1}(X^TX + 2Z^{*T}X + Z^{*T}Z^*)$$

$$(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$$

$$= I(X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$$

$$= (X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2 \quad (3.16)$$

Sehingga $Var(\hat{\beta}^*) = (X^TX + 2X^TZ^* + Z^{*T}Z^*)^{-1}\sigma^2$ harus sekecil mungkin agar $\hat{\beta}^*$ efisien.

3.3.3 Konsisten

Estimator yang konsisten adalah

$$E(\hat{\theta} - E(\theta))^2 \to 0 \text{ jika } n \to \infty$$

Sehingga

$$E\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)^2 = E\left[(\hat{\beta}^* - E(\hat{\beta}^*))(\hat{\beta}^* - E(\hat{\beta}^*))^T\right]$$

Dari persamaan (3.15) diperoleh $E(\hat{\beta}^*) = \beta^*$ maka

$$E\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)^2 = E\left[(\hat{\beta}^* - E(\hat{\beta}^*))(\hat{\beta}^* - E(\hat{\beta}^*))^T\right]$$

$$= E\left[(\hat{\beta}^* - \beta^*)(\hat{\beta}^* - \beta^*)^T\right]$$

$$= E(\hat{\beta}^* - \beta^*)(\hat{\beta}^* - \beta^*)^T$$

$$= \left(E(\hat{\beta}^*) - E(\beta^*)\right)(\hat{\beta}^* - \beta^*)^T$$

$$= (\beta^* - \beta^*)(\hat{\beta}^* - \beta^*)^T$$

$$= (0)(\hat{\beta}^* - \beta^*)^T$$

$$= 0 \tag{3.17}$$

Dari persamaan (3.17) diperoleh $E\left[(\hat{\beta}^* - E(\hat{\beta}^*))(\hat{\beta}^* - E(\hat{\beta}^*))^T\right] = 0$, maka untuk $\hat{\beta}^*$ yang mengandung *outlier* merupakan estimator yang konsisten.

3.4 Aplikasi Pada Estimasi Parameter Model Regresi Linier yang Mengandung *Outlier*

3.4.1 Diskripsi Data

Data berikut berasal dari Neil H. Timm (1975: 281), data ini adalah data 32 siswa dari sekolah bangsa kulit putih kelas atas (*upper-class*) yang dipilih secara acak. Data ini merupakan data tes kemampuan berbahasa Inggris. Datanya adalah sebagai berikut:

Tabel 3.1 Tes Kemampuan Berbahasa Inggris

No	Y	X_1	X_2	X_3	X_4
1	68	0	10	21	22
2	82	7	3	28	21
3	82	7	9	31	30
4	91	6	11	27	25
5	82	20	7	28	16
6	100	4	11	32	29
7	100	6	7	26	23
8	96	5	2	22	23
9	63	3	5	24	20
10	91	16	12	27	30

11	87	5	3	25	24
12	105	2	11	26	22
13	87	1	4	25	19
14	76	11	5	27	22
15	66	0	0	16	11
16	74	5	8	12	15
17	68	1	6	28	23
18	98	1	9	30	18
19	63	0	13	19	16
20	94	4	6	27	19
21	82	4	5	21	24
22	89	1	6	23	28
23	80	5	8	25	24
24	61	4	5	16	22
25	102	5	7	26	15
26	71	0	4	16	14
27	102	4	17	27	31
28	96	5	8	28	26
29	55	4	7	20	13
30	96	4	7	23	19
31	74	2	6	25	17
32	78	5	10	27	26

Sumber: Neil H. Timm, 1975. Multivariate Analysis With Applications in Education Psychologi. Publishing Company. Monterey California.

keterangan:

Y : tes kemampuan berbahasa Inggris

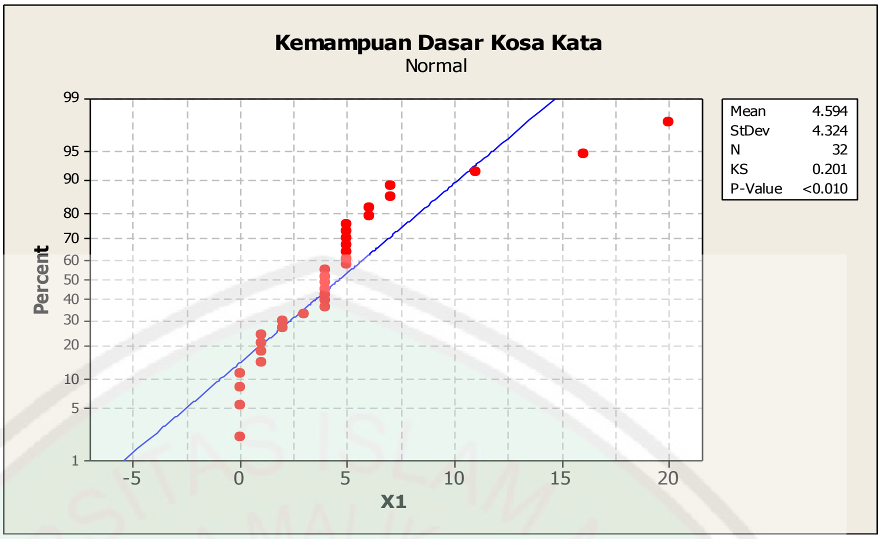
X_1 : kemampuan dasar kosa kata (*vocab*)

X_2 : kemampuan membaca

X_3 : kemampaun praktik

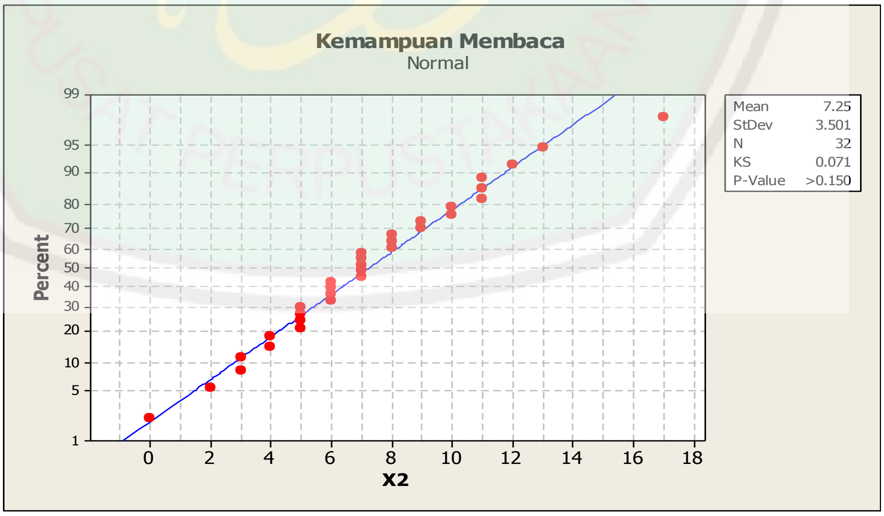
X_4 : kemampuan *grammar*

Dari data Tabel 3.1 dapat dibuat grafik dengan menggunakan MINITAB 14 dan mengasumsikan $\alpha = 0{,}05$. Hasilnya adalah sebagai berikut:



Gambar 3.1 Kemampuan Dasar Kosa Kata

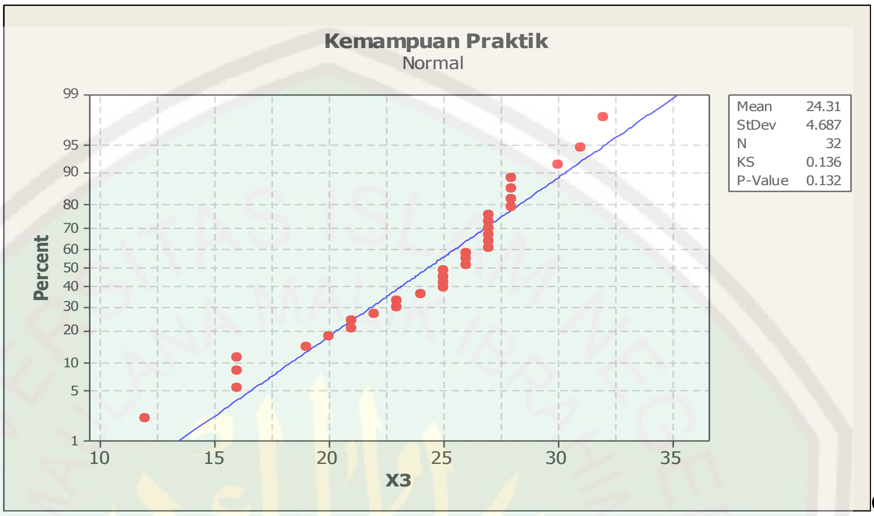
Gambar 3.1 di atas merupakan gambar kemampuan dasar kosa kata (X_1) dari 32 siswa, dengan nilai rata-rata sebesar 4,594 dan standar deviasi 4,324. Dari gambar 3.1 juga dapat dilihat bahwa sebaran data tidak normal, karena nilai p-value $< \alpha$. Jika dilihat sekilas ada beberapa data yang letaknya jauh dari garis normal.



Gambar 3.2 Kemampuan Membaca

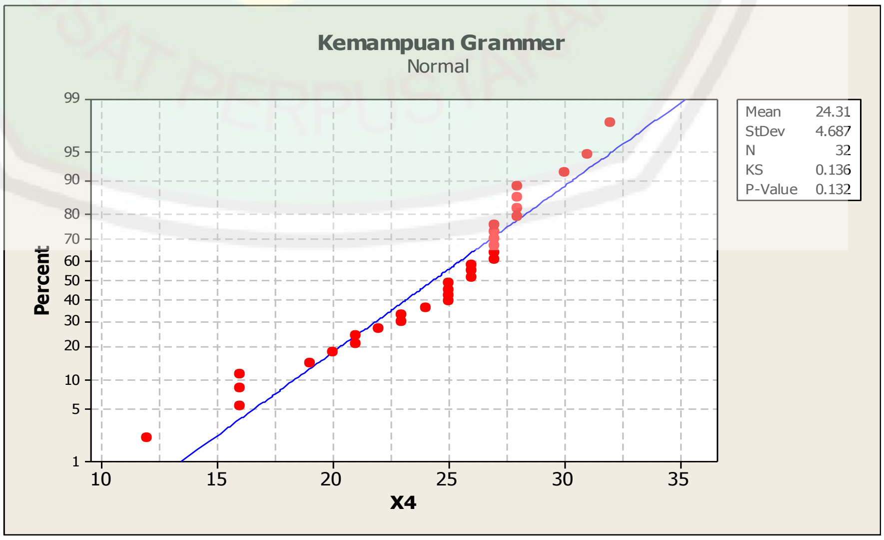
Gambar 3.2 di atas merupakan gambar kemampuan membaca (X_2) dari 32 siswa, dengan nilai rata-rata sebesar 7,25 dan standar deviasi

3,501. Dari gambar 3.2 jika dilihat sekilas ada satu data yang letaknya jauh dari garis normal, akan tetapi sebaran data itu normal, karena nilai p-value > α.



Gambar 3.3 Kemampuan Praktik

Gambar 3.3 di atas merupakan gambar kemampuan praktik (X_3) dari 32 siswa, dengan nilai rata-rata sebesar 24,31 dan standar deviasi 4,687. Dari gambar 3.3 juga dapat dilihat bahwa sebaran data normal, karena nilai p-value > α.



Gambar 3.4 Kemampuan *Grammar*

Gambar 3.4 di atas merupakan gambar kemampuan *grammar* (X_4) dari 32 siswa, dengan nilai rata-rata sebesar 24,31 dan standar deviasi 4,687. Dari gambar 3.4 juga dapat dilihat bahwa sebaran data normal, karena nilai p-value > α.

3.4.2 Analisis Data pada Model yang Mengandung *Outlier*

Setelah didapat model dari data pada tabel 3.1, selanjutnya dilakukan analisis data dengan menggunakan MINITAB 14. Dengan hipotesis sebagai berikut:

Hipotesis untuk uji F:

H_0: ada salah satu parameter model yang bernilai nol

H_1: tidak ada parameter model yang bernilai nol

Hipotesis untuk uji t:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4$$

$$H_1: \beta_0 \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Untuk hipotesis awal berarti tidak ada pengaruh antara kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar* terhadap tes kemampuan berbahasa Inggris. Sedangkan hipotesis alternatifnya ada pengaruh antara kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar* terhadap tes kemampuan berbahasa Inggris.

Sehingga didapat model analisis regresi dari pengaruh antara kemampuan dasar kosa kata (X_1), kemampuaan membaca (X_2), kemampuan praktik (X_3), kemampuan *grammar* (X_4), terhadap tes kemampuan berbahasa Inggris (Y), adalah:

$$Y = 38{,}9 - 0{,}084\ X_1 + 0{,}318\ X_2 + 1{,}40X_3 + 0{,}383X_4 \qquad (3.1)$$

Dari persamaan (3.1) didapatkan β_0 sebesar 38,9, estimasi parameter β_1 sebesar -0,08, estimasi parameter β_2 sebesar 0,318, estimasi parameter β_3 sebesar 1,40 dan estimasi parameter β_4 sebesar 0,3832. Jika dari masing-masing X (kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar*) nilainya meningkat maka akan meningkat pula nilai Y (tes kemampuan berbahasa Inggris).

Setelah didapatkan model untuk selanjutnya dilakukan uji ANOVA. Tujuannya untuk memeriksa apakah X (kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik, kemampuan *grammar*) secara serentak mempunyai pengaruh terhadap Y (tes kemampuan berbahasa Inggris), diperoleh tabel sebgai berikut:

Tabel 3.2 Anova Pada Tes Kemampuan Berbahasa Inggris yang Mengandung *Outlier*

Source	DF	SS	MS	F	P
Regression	4	2095,0	523,8	3,59	0,018
Residual Error	27	3937,7	145,8		
Total	31	6032,7			

Dari tabel ANOVA 3.2 diperoleh uji F sebesar 3,59, dengan menggunakan α sebesar 5% maka $F_{4,27}^{0,05}$ = 2,73, sehingga $F_{hitung} > F_{tabel}$ maka H_0 ditolak atau ada pengaruh antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris).

Dari uji ANOVA tersebut telah dikatakan bahwa terdapat pengaruh antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris), akan tetapi tidak diketahui antara kemampuan dasar kosa kata (X_1), kemampuaan membaca (X_2), kemampuan praktik (X_3), dan kemampuan grammer (X_4), yang mana, yang berpengaruh terhadap tes kemampuan berbahasa Inggris (Y). Untuk mengetahui hal tersebut, maka perlu dilakukan uji t, seperti pada tabel berikut:

Tabel 3.3 Estimasi Parameter β yang Mengandung *Outlier*

	Coef	SE Coef	T	P
β_0	38,95	12,01	3,24	0,003
β_1	-0,0840	0,5341	-0,16	0,876
β_2	0,3184	0,6985	0,46	0,652
β_3	1,3984	0,5813	2,41	0,023
β_4	0,3832	0,5340	0,72	0,479

Dari tabel 3.3 didapatkan β_0 sebesar 38,95, estimasi parameter β_1 sebesar -0,0840, estimasi parameter β_2 sebesar 0,3184, estimasi parameter β_3 sebesar 1,3984 dan estimasi parameter β_4 sebesar 0,3832.

Selain itu, tabel 3.3 juga memperlihatkan nilai dari t_{hitung} dari masing-masing β yang mana t_{hitung} ini akan dibandingkan dengan t_{tabel}. Dengan menggunakan $\alpha = 5\% = 0{,}05$ maka t_{tabel} adalah $t_{31}^{005} = 1{,}697$ kemudian t_{tabel} dibandingkan dengan t_{hitung}. Untuk $\beta_1, t_{hitung} < t_{tabel}$, yaitu $0{,}16 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh antara kemampuan dasar kosa kata (X_1) terhadap tes kemampuan berbahasa Inggris (Y). Untuk β_2, $t_{hitung} < t_{tabel}$ yaitu $0{,}46 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh antara kemampuan membaca (X_2) terhadap tes kemampuan berbahasa Inggris (Y). Untuk β_3, $t_{hitung} > t_{tabel}$, yaitu $2{,}41 > 1{,}697$ maka H_0 ditolak atau ada pengaruh antara kemampuan praktik (X_3) terhadap tes kemampuan berbahasa Inggris (Y). Untuk β_4, $t_{hitung} < t_{tabel}$, yaitu $0{,}72 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh kemampuan *grammar* (X_4) terhadap tes kemampuan berbahasa Inggris (Y).

Dalam regresi ada istilah *Mean Square Error* (MSE) yang merupakan varian error (s^2). Varian error adalah kuadrat standar deviasi. Nilai MSE untuk model yang telah dibuat adalah 145,8. Jadi nilai standar deviasi model adalah:

$$s = \sqrt{145{,}8} = 12{,}07$$

Semakin kecil standar errornya berarti nilai estimasi model semakin mendekati nilai sebenarnya.

Untuk mengetahui besarnya pengaruh antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris), dapat dilihat nilai koefisien determinasi (r^2). Nilai koefisien determinasi yang diperoleh adalah 34,7%. Koefisien korelasi r merupakan akar determinasi, nilai r ini berkisar antara 0 sampai 1, di mana semakin mendekati 1 berarti hubungan antar variabel makin kuat. Dalam hal ini koefisien korelasinya sebesar;

$$r = \sqrt{0{,}347} = 0{,}589$$

Makin banyak variabel yang dimasukkan dalam model maka makin meningkatkan nilai r^2. Padahal dengan semakin banyak variabel, model menjadi tidak efisien.

Antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar)* terhadap Y (tes kemampuan berbahasa Inggris), dapat dilihat nilai koefisien korelasinya sebagai berikut:

Tabel 3.4 Korelasi PadaTes Kemampuan Berbahasa Inggris yang Mengandung *Outlier*

	Y	X_1	X_2	X_3	X_4
Y	1	0,174	0,290	0,566	0,436
X_1	0,174	1	0,105	0,342	0,241
X_2	0,290	0,105	1	0,314	0,456
X_3	0,566	0,342	0,314	1	0,556
X_4	0,436	0,241	0,456	0,556	1

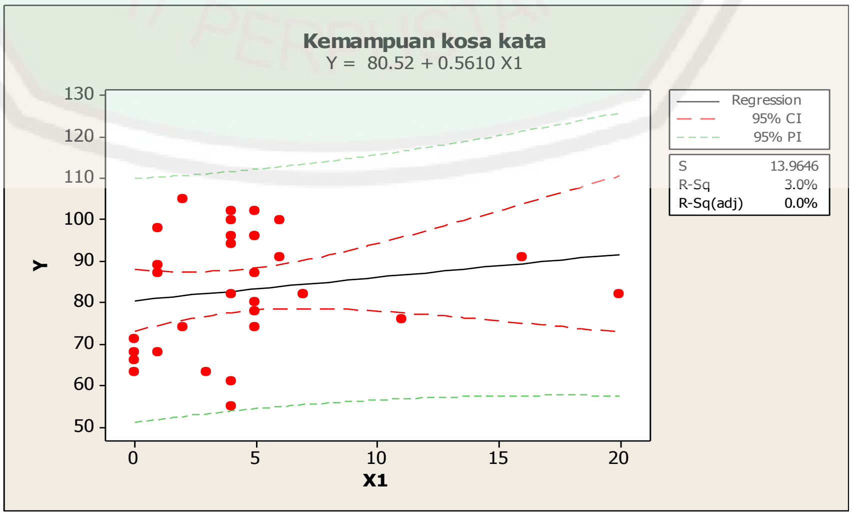
Dari tabel 3.4 dapat dilihat bahwa yang mempunyai korelasi yang kuat adalah korelasi antara variabel itu sendiri misalnya, korelasi X_1 dengan X_1 atau korelasi kemampuan kosa kata terhadap kemampuan kosa kata itu sendiri, korelasinya sebesar 1. Korelasi antara kemampuan dasar kosa kata (X_1) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,174, hal ini bisa dikatakan korelasinya lemah karena nilai korelasi jauh dari 1. Korelasi antara kemampuan membaca (X_2) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,290, hal ini bisa dikatakan korelasinya lemah karena nilai korelasi jauh dari 1. Korelasi antara kemampuan praktik (X_3) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,566, hal ini bisa dikatakan korelasinya lumayan kuat karena nilai korelasi sudah mendekati 1 meskipun agak jauh, begitu juga korelasi antara kemampuan *grammar* (X_4) terhadap tes kamampuan berbahasa Inggris (Y) lumayan kuat, nilai korelasinya sebesar 0,436.

Hasil analisis di atas adalah hasil analisis pada data yang masih mengandung *outlier*. Hasil yang terdapat *outlier* dapat dilihat pada tabel berikut:

Tabel 3.5 *Outlier* Pada Tes Kemampuan Berbahasa Inggris

Obs	X_1	Y	Fit	SE Fit	SE Resid
5	20,5	82,00	84,78	-2,78	-0,35

Dari tabel 3.5 bisa diketahui data yang merupakan *outlier* adalah data ke 5 dari variabel bebas X_1 atau kemampuan dasar kosa kata. Karena data ke-5 diduga data yang *outlier* maka perlu ditunjukkan kebenaran *outlier*nya dengan membuat *fit line plot*.



Gambar 3.5 *Outlier* pada Kemampuan Dasar Kosa Kata

Dari gambar 3.5 di atas terlihat jelas bahwa ada 1 data yang berada jauh dari sekumpulan data yang lain, akan tetapi data itu tidak keluar dari garis selang kepercayaan (pada garis warna hijau). Data itu adalah data ke-5, sehingga data ke-5 ini merupakan data *outlier*, karena data ke-5 ini adalah data *outlier* maka perlu dihilangkan keberadaanya, karena biasanya *oulier* inilah yang menyebabkan data tidak menyebar normal dan error yang diperoleh besar. Selanjutnya di bawah ini akan dilanjutkan analisis data yang tidak mengandung *outlier* yaitu dengan membuang data ke-5.

3.4.3 Analisis Data Untuk Model *Outlier* Dihilangkan

Setelah diketahui bahwa terdapat *outlier* pada X_1 data ke-5, maka selanjutnya data tersebut dihilangkan kemudian dilakukan analisis terhadap model *outlier* yang dihilangkan tersebut dengan langkah-langkah seperti langkah-langkah pada analisis data yang mengandung *outlier*.

Hipotesis untuk uji F:

H_0: ada salah satu parameter model yang bernilai nol

H_1: tidak ada parameter model yang bernilai nol

Hipotesis untuk uji t:

$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4$

$$H_1: \beta_0 \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

Untuk hipotesis awal berarti tidak ada pengaruh antara kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar* terhadap tes kemampuan berbahasa Inggris. Sedangkan hipotesis alternatifnya ada pengaruh antara kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar* terhadap tes kemampuan berbahasa Inggris.

Sehingga didapat model analisis regresi dari pengaruh antara kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik, kemampuan *grammar*, terhadap tes kemampuan berbahasa Inggris adalah

$$Y = 39{,}4 - 0{,}098\ X_1 + 0{,}343\ X_2 + 1{,}43X_3 + 0{,}286X_4 \qquad (3.2)$$

Dari persamaan (3.2) didapatkan β_0 sebesar 39,4 dan estimasi parameter β_1 sebesar 0,098, estimasi parameter β_2 sebesar 0,343, estimasi parameter β_3 sebesar 1,43 dan estimasi parameter β_4 sebesar 0,286. Jika dari masing-masing X (kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik dan kemampuan *grammar*) nilainya meningkat maka akan meningkat pula nilai Y (tes kemampuan berbahasa Inggris)

Setelah didapatkan model untuk selanjutnya dilakukan uji ANOVA. Tujuannya untuk memeriksa apakah X (kemampuan dasar kosa kata, kemampuaan membaca, kemampuan praktik, kemampuan *grammar*) secara serentak mempunyai pengaruh terhadap Y (tes kemampuan berbahasa Inggris), diperoleh tabel sebgai berikut:

Tabel 3.6 Anova Pada Tes Kemampuan Berbahasa Inggris Ketika *Outlier* Dihilangkan

Source	DF	SS	MS	F	P
Regression	4	2111,7	527,9	3,50	0,020
Residual Error	26	3919,8	150,8		
Total	30	6031,5			

Dari tabel ANOVA 3.6 diperoleh uji F sebesar 3,50, dengan menggunakan α sebesar 5% maka $F_{4,26}^{0,05}$ = 2,74, sehingga $F_{hitung} > F_{tabel}$ maka H_0 ditolak atau ada pengaruh antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris).

Dari uji ANOVA tersebut telah dikatakan bahwa terdapat pengaruh antara X(kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris), akan tetapi tidak diketahui antara kemampuan dasar kosa kata (X_1), kemampuaan membaca (X_2), kemampuan praktik(X_3), dan kemampuan *grammar* (X_4), yang mana, yang berpengaruh terhadap tes kemampuan berbahasa Inggris (Y). Untuk mengetahui hal tersebut, maka perlu dilakukan uji t, seperti pada tabel berikut:

Tabel 3.7 Estimasi Parameter β Ketika *Outlier* Dihilangkan

	Coef	SE Coef	T	P
β_0	39,44	12,30	3,21	0,004
β_1	0,0979	0,7578	0,13	0,898
β_2	0,3434	0,7139	0,48	0,635
β_3	1,4301	0,5982	2,39	0,024
β_4	0,2863	0,6115	0,47	0,644

Dari tabel 3.7 didapatkan β_0 sebesar 39,44, estimasi parameter β_1 sebesar 0,09779, estimasi parameter β_2 sebesar 0,3434, estimasi parameter β_3 sebesar 1,4301 dan estimasi parameter β_4 sebesar 0,2863.

Selain itu, tabel 3.7 juga memperlihatkan nilai dari t_{hitung} dari masing-masing β yang mana t_{hitung} ini akan dibandingkan dengan t_{tabel}.

Dengan menggunakan $\alpha = 5\% = 0{,}05$ maka t_{tabel} adalah $t_{30}^{005} = 1{,}697$ kemudian t_{tabel} dibandingkan dengan t_{hitung}. Untuk $\beta_1, t_{hitung} < t_{tabel}$, yaitu $0{,}13 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh antara kemampuan dasar kosa kata (X_1) terhadap tes kemampuan berbahasa Inggris (Y). Untuk β_2, $t_{hitung} < t_{tabel}$ yaitu $0{,}48 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh antara kemampuan membaca (X_2) terhadap tes kemampuan berbahasa Inggris (Y). Untuk $\beta_3, t_{hitung} > t_{tabel}$, yaitu $2{,}39 > 1{,}697$ maka H_0 ditolak atau ada pengaruh antara kemampuan praktik (X_3) terhadap tes kemampuan berbahasa Inggris (Y). Untuk β_4, $t_{hitung} < t_{tabel}$, yaitu $0{,}47 < 1{,}697$ maka H_0 diterima atau tidak ada pengaruh kemampuan *grammar* (X_4) terhadap tes kemampuan berbahasa Inggris (Y).

Dalam regresi ada istilah *Mean Square Error* (MSE) yang merupakan varian error (s^2). Varian error adalah kuadrat standar deviasi. Nilai MSE untuk model yang telah dibuat adalah 150,8. Jadi nilai standar deviasi model adalah:

$$s = \sqrt{150{,}8} = 12{,}28$$

Untuk mengetahui besarnya pengaruh antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris), dapat dilihat nilai koefisien determinasi (r^2). Nilai koefisien determinasi yang diperoleh adalah 35,7%. Koefisien korelasi r merupakan akar determinasi, nilai r ini berkisar antara 0 sampai 1, di mana semakin mendekati 1 berarti hubungan antar variabel makin kuat. Dalam hal ini koefisien korelasinya sebesar;

$$r = \sqrt{0{,}35} = 0{,}591$$

Antara X (kemampuan dasar kosa kata, kemampuan membaca kemampuan praktik dan kemampuan *grammar*) terhadap Y (tes kemampuan berbahasa Inggris), dapat dilihat nilai koefisien korelasinya sebagai berikut:

Tabel 3.8 Korelasi Tes Kemampuan Berbahasa Inggris Ketika *Outlier* Dihilangkan

	Y	X_1	X_2	X_3	X_4
Y	1	0,241	0,290	0,574	0,442
X_1	0,241	1	0,149	0,331	0,453
X_2	0,290	0,149	1	0,319	0,462
X_3	0,574	0,331	0,319	1	0,061
X_4	0,442	0,453	0,462	0,061	1

Dari tabel 3.4 dapat dilihat bahwa yang mempunyai korelasi yang kuat adalah korelasi antara variabel itu sendiri misalnya, korelasi X_2 dengan X_2 atau korelasi kemampuan membaca terhadap kemampuan membaca itu sendiri, korelasinya sebesar 1. Korelasi antara kemampuan kosa kata (X_1) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,241, hal ini bisa dikatakan korelasinya lemah karena nilai korelasi jauh dari 1. Korelasi antara kemampuan membaca (X_2) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,290, hal ini bisa dikatakan korelasinya lemah karena nilai korelasi jauh dari 1. Korelasi antara kemampuan praktik (X_3) terhadap tes kemampuan berbahasa Inggris (Y) sebesar 0,574, hal ini bisa dikatakan korelasinya agak kuat karena nilai korelasi sudah mendekati 1 meskipun agak jauh, begitu juga korelasi antara kemampuan *grammar* (X_4) terhadap tes kamampuan berbahasa Inggris (Y) agak kuat, nilai korelasinya sebesar 0,442.

Hasil analisis di atas adalah hasil analisis data yang *outlier*nya dihilanghan. Dari hasil analisis pada data yang mengandung *outlier* dan analisis data ketika *oulier*nya dihilangkan bisa dilihat perbandingannya yaitu nilai estimasi parameter model regresi linier pada data yang mengandung *outlier* lebih kecil daripada nilai estimasi parameter model regresi linier pada data ketika *outlier*nya dihilangkan. Jika dilihat dari standar errornya, standar error pada data yang mengandung *outlier* lebih kecil dari pada standar error pada data ketika *outlier*nya.

Padahal secara teori, kenyataannya jika terdapat *outlier* seharusnya estimasi yang didapat harus lebih besar dibanding dengan yang tidak mengandung *outlier*. Begitu juga dengan standar error yang mengandung *outlier* harus lebih besar daripada standar error yang tidak yang mengandung *outlier*.

Pada penelitian ini hasilnya justru berkebalikan, hal ini berarti bahwa *outlier* yang berada pada data tidak boleh dibuang, karena *outlier* tersebut keberadaannya sangat berpengaruh. Keberadaan *outlier* tersebut perlu diatasi agar tidak mengganggu (merusak) data yang lain salah satunya mungkin dengan transformasi data. Di sini peneliti tidak melanjutkan pada cara-cara mengatasi *outlier* karena hal ini bukan tugas peneliti.

3.3 Keterkaitan Hasil Penelitian dengan Kajian Agama

Pada BAB II telah disinggung bahwa estimasi parameter terdapat pada surat Ash-Shaffaat ayat 147. Peneliti pada bab ini akan menghubungkan antara Qs. Ash-Shaffaat ayat 147 dengan konsep estimasi dalam matematika. Konsep estimasi dalam matematika ternyata telah terkonsep sejak zaman Nabi Muhammad SAW. Hal tersebut terbukti dijelaskan dalam Al-Qur'an surat Ash-Shaffaat ayat 147, yang secara tidak langsung telah melahirkan konsep pendugaan.

وَأَرْسَلْنَٰهُ إِلَىٰ مِاْئَةِ أَلْفٍ أَوْ يَزِيدُونَ ١٤٧

Artinya: "*Dan Kami utus Dia kepada seratus ribu orang atau lebih*" (Qs. Ash-Shaaffat 37:147*).*

Pengertian pendugaan dalam surat Ash-Shaffaat ayat 147 merupakan estimasi banyak, maksudnya menghitung jumlah umat Nabi Yunus tidak secara eksak, yaitu melalui penaksiran. Dari sini diketahui bahwa estimasi dalam ayat tersebut merupakan pendugaan dalam konsep yang sederhana dan dalam matematika digunakan untuk perhitungan-perhitungan dasar matematika.

Kaitan estimasi pada surat ini terletak pada kalimat "*seratus ribu orang atau lebih*", kalimat tersebut menjelaskan dalam menentukan jumlah umat Nabi Yunus tidak dengan perhitungan secara eksak. Sehingga terdapat perbedaan pendapat para ulama' dalam menafsirkan ayat tersebut. Jika dipahami dalam arti "atau", maka ayat ini menyatakan jumlah mereka banyak, menurut perhitungannnya adalah seratus ribu/lebih. Jika dipahami dalam arti "dan/bahkan", maka artinya adalah mereka diutus kepada dua kelompok, yang pertama berjumlah seratus ribu (100.000) dan yang satu lagi adalah yang lebih dari itu.

Al-Maraghi dalam Tafsir Al-Maraghi (1974:138), menceritakan bahwa Nabi Yunus diutus oleh kaum itu dan mereka ada 100.000 bahkan lebih. Maka menjadi stabil keadaan mereka dan beriman kepada Yunus. Karena, setelah Yunus keluar dari kalangan mereka, mereka berpikir benar-benar telah melakukan kekeliruan, dan jika mereka tidak mengikuti Rasul, maka mereka akan binasa, seperti yang terjadi atas umat-umat sebelum mereka. Maka tatkala Yunus kembali kepada mereka dan menyeru kepada Tuhannya, maka mereka menyambut seruan Yunus itu dengan taat dan tunduk kepada perintah dan larangan Allah. Maka kami anugrahi kenikmatan kepada mereka dalam kehidupan ini hingga ajal, dan mereka pun mati sebagaimana matinya orang-orang lain.

Para ulama’ diatas mempunyai versi yang berbeda-beda dalam menafsirkan kalimat “*kepada* seratus *ribu orang atau lebih*” karena ayat tersebut tidak ada kejelasan dalam menerangkan jumlah umat Nabi Yunus. Para ulama memperkirakan jumlah umat Nabi Yunus dengan jumlah yang berbeda-beda tetapi meskipun demikian tidak ada yang mengatakan kurang dari 100.000 orang.

Dari ayat diatas diketahui bahwa terdapat perbedaan pendapat para ulama’ dalam menduga banyaknya umat Nabi Yunus. Kata yang bermakna *lebih* itu oleh para ulama’ diduga sebanyak 20.000 orang, 30.000 orang, atau 70.000 orang. Ada juga yang hanya mengatakan *lebih saja*. Jika mengatakan *lebih saja*, maka bisa saja 10.000 orang atau 15.000 orang, hal ini karena ayat tersebut tidak mengatakan jumlah umat Nabi Yunus yang sebenarnya. Jika umat Nabi Yunus dapat dinyatakan dalam X, maka niali X tersebut berada dalam skala interval100.000 $\leq X <$ 200.000, artinya umat Nabi Yunus tidak kurang dari 100.000 dan tidak sampai 200.000 orang.

Beberapa penjelasan ayat Al-Qur'an di atas menggambarkan dengan jelas bahwa teori ekspektasi dan estimasi sudah ada sekitar 1400 tahun yang lalu, akan tetapi teori tersebut dipelajari secara mendalam sekitar tahun 1980-an di Amerika. Sehingga dalam kehidupan sehari-hari, ketrampilan estimasi sangat dibutuhkan dan menghemat waktu dalam suatu penghitungan.

Perbedaan estimasi dalam surat Ash-Shaffaat dengan estimasi parameter dalam penelitian ini terletak pada objek yang diestimasi dan syarat atau sifat-sifat yang harus dipenuhi. Dalam surat Ash-Shaffaat ayat 147 menduga terhadap banyaknya jumlah dan syarat penduga berupa interval yaitu 100.000 $\leq X <$ 200.000, sedangkan dalam penelitian ini mengestimasi model regresi yang estimatornya berupa rumus, yang dapat diterapkan dalam penelitian. Estimasi tersebut dengan syarat harus memenuhi sifat-sifat yaitu unbias, konsisten dan efisien. Dari sini perlu diketahui, bahwa ilmu pengetahuan umum seperti matematika khususnya konsep estimasi parameter, yang diyakini oleh sebagian orang diciptakan oleh orang-orang barat nonmuslim, ternyata telah terkonsep dalam Al-Qur'an. Hal ini membuktikan bahwa Al-Qur'an tidak hanya berbicara tentang ilmu-ilmu agama saja, akan tetapi juga berbicara tentang ilmu pengetahuan umum. Dalam Al-Qur'an, konsep-konsep ilmu pengetahuan tidak disajikan secara langsung, akan tetapi berupa pengetahuan yang membutuhkan penafsiran secara mendalam.

Metode estimasi juga disebutkan dalam suatu hadits pada bab jual-beli yang diriwayatkan Malik, dari Nafi’, dari Abdullah bin Umar

عَنْ مَا لِكٍ عَنْ نَا فِعٍ عَن عَبدِ اللهِ بْنِ عُمَرَ رَضِيَ الله عَنهُمَا أَنّ رَسُولُ اللهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ نَهَى عَنِ اْ لمُزَا بَنَةِ, وَاْ لمُزَا بَنَةُ اشْتِرَاءُ الثَّمَرِ بِا لتَّمْرِ كَيْلاً وَبَيْعُ ا لكَرْمِ بِا لزَ بِيْبِ كَيْلاً

“*Dari Malik, dari Nafi’, dari Abdullah bin Umar ra bahwa Rasulullah SAW melarang jual beli muzabanah. Muzabanah aalah membeli kurma basah [yang masih berada di atas pohon] dengan kurma kering berdasarkan takaran, dan menjual anggur yang basah dengan anggur kering berdasarkan takaran*” (Shahih Bukhari, 2185).

Muzabanah yaitu menjual buah berdasarkan takaran, kaitanya dengan estimasi bahwasanya buah tersebut dijual dengan cara di taksir atau dengan memperkirakan. Dalam artian buah kurma yang masih berada di atas pohon dijual dengan cara ditaksir (dikira-kira) seharga atau sejumlah dengan buah kurma kering dan begitu juga dengan buah anggur basah dijual dengan cara ditaksir (dikira-kira) seharga atau sejumlah dengan buah anggur kering.

Selain estimasi, dalam penelitian ini juga menyinggung tentang *outlier*, yang mana dalam Al-Qur’an telah dijelaskan dalam surat Al-Jin ayat 14 sebagai berikut:

وَأَنَّا مِنَّا ٱلْمُسْلِمُونَ وَمِنَّا ٱلْقَٰسِطُونَ ۖ فَمَنْ أَسْلَمَ فَأُو۟لَٰٓئِكَ تَحَرَّوْا۟ رَشَدًا ﴿١٤﴾

Artinya: “*Dan Sesungguhnya di antara Kami ada orang-orang yang taat dan ada (pula) orang-orang yang menyimpang dari kebenaran.*

Barangsiapa yang yang taat, maka mereka itu benar-benar telah memilih jalan yang lurus" (QS. Al-Jin 72:14*).*

Surat Al-Jin ayat 14 di atas, Allah menjelaskan tentang jin bahwa di antara mereka ada yang beriman mentaati Allah, khusyuk dan ikhlas serta beramal saleh karena-Nya. Ada pula di antara mereka yang berpaling dari haluan yang benar. Barangsiapa beriman kepada Allah dan mentaati-Nya sesungguhnya dia telah menempuh jalan yang akan menyampaikannya kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkannya dari siksa neraka.

Setelah diuraikan di atas dapat diambil kesimpulan bahwa yang menjelaskan *outlier* adalah kalimat "*Dan sesungguhnya di antara kami ada orang-orang yang taat dan ada pula orang-orang yang menyimpang dari kebenaran*" dalam artian *outlier* adalah suatu penyimpangan.

Dari penafsiran ayat ini dijelaskan bahwa para penyimpang yakni mereka yang telah sangat jauh dari kebenaran lagi sangat mantap kekufurannya. Penyimpangan ini mempunyai arti yang sama dengan *outlier* yaitu sama-sama terletak sangat jauh diantara data dalam model.

BAB IV

PENUTUP

4.1 Kesimpulan

Berdasarkan pembahasan dari bab tiga, maka dapat disimpulkan bahwa estimasi parameter pada model regresi linier yang mengandung *outlier* dengan metode *maximum likelihood estimation* menghasilkan suatu estimasi parameter yaitu:

$$\hat{\beta}^* = \left(X^T X + 2Z^{*T} X + Z^{*T} Z^*\right)^{-1}\left(X^T + Z^{*T}\right)Y^*$$

$$\hat{\sigma}^2 = \frac{1}{n}\left(Y^* - (X\beta^* + Z^*\beta^*)\right)^T\left(Y^* - (X\beta^* + Z^*\beta^*)\right)$$

Sifat-sifat model regresi linier yang mengandung *outlier* memenuhi sifat-sifat dari estimasi parameter yang baik yaitu unbias, efisien dan konsisten. Nilai estimasi parameter yang dihasilkan model regresi linier pada data yang mengandung *outlier* lebih kecil daripada nilai estimasi parameter model regresi linier pada data ketika *outlier*nya dihilangkan.

4.2 Saran

Dalam penelitian ini peneliti menggunakan model regresi linier yang mengandung *outlier* dan mengaplikasikan langsung pada data yang mengandung *outlier* dengan metode *maximum likelihood estimation*. Bagi pembaca yang ingin melakukan penelitian serupa, peneliti menyarankan menggunakan metode yang lain dengan model regresi linier yang mengandung *outlier*.

DAFTAR PUSTAKA

Abdusysyakir. 2007. *Ketika Kyai Mengajar Matematika*. Malang: UIN Malang Press

Al-Maraghi, Mushthafa Ahmad. 1989. *Tafsir Al-Maraghi Jilid 25*. Semarang: CV. Toha Putra Semarang

Draper, Norman dan Harry Smith, 1992. *Analisis Regresi Terapan*. PT Gramedia Pustaka Utama: Jakarta

Dudewich & Mishra. 1995. *Statistik Matematika Modern*. Bandung: ITB.

Gujarati, N. Damodar. 2007. *Dasar-Dasar Ekonometrika Jilid 1*. Jakarta: Erlangga

Hasan, Iqbal. 2002. *Pokok-Pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta: Bumi Aksara

Hasan, Iqbal. 2002. *Pakok-Pokok Materi Metodologi Penelitian dan Aplikasinya*. Jakarta: Ghalia Indonesia

Mardalis. I990. *Metode Penelitian Suatu Pendekatan Proposal*. Jakarta: Bumi Aksara

Mood, M. Alexander dkk. 1986. *Introduction To The Theory Of Statistics*. Megarw Hill Book Company

Murray dan Larry. 2007. *Statistik Edisi Ke 3*. Jakarta: Erlangga

Soemarti. 2007. *Pencilan (Outlier)*. Makalah Statistika FMIPA Universitas Padjadjaran, Bandung. Tersedia: http:///resources.unpad.ac.id/unpad-content/uploads/publikasi_dosen/Outlier(Pencilan).pdf (diunduh pada tanggal 15 Oktober 2011)

Supangat, Andi. 2008. *Statistik dalam Kajian Deskriptif, Inferensia dan Nonparametrik.* Jakarta: Kencana

Walpole, E. Ronald dan Myers, Raymond. 1995. *Ilmu Peluang dan Statistik untuk Insinyur* dan *Ilmuwan*. Bandung: ITB Bandung

Wibisono, Yusuf. 2005. *Metode Statistik*. Yogyakarta: Gajah Mada University Press

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

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI (UIN)
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No. 50 Dinoyo Malang (0341) 551345 fax (0341) 572533

BUKTI KONSULTASI SKRIPSI

Nama : **Nur Ngaini**
NIM : **08610072**
Fakultas/ Jurusan : **Sains dan Teknologi/ Matematika**
Judul : **Estimasi Parameter Model Regresi Linier pada Data yang Mengandung *Outlier* dengan Metode *Maximum Likelihood Estimation***
Dosen pembimbing I : **Sri Harini, M.Si**
Dosen pembimbing II : **Dr. Ahmad Barizi, MA**

No.	Tanggal	Hal	Tanda tangan	
1	18 September 2011	Proposal Skripsi	1	
2	24 September 2011	ACC Proposal Skripsi		2
3	19 November 2011	Konsultasi BAB I	3	
4	26 November 2011	Konsultasi BAB I dan BAB II		4
5	08 Desember 2011	Konsultasi BAB I dan BAB II Keagamaan	5	
6	09 Desember 2011	Konsultasi BAB I, BAB II dan BAB III		6
7	12 Desember 2011	Konsultasi BAB I, BAB II dan BAB III	7	
8	21 Desember 2011	ACC BAB I dan BAB II		8
9	02 Januari 2012	Konsultasi BAB III	9	
10	07 Januari 2011	Konsultasi BAB III		10
11	12 Januari 2012	ACC BAB III	11	
12	12 Januari 2012	Konsultasi BAB III		12
13	12 Januari 2012	ACC Keagamaan BAB III	13	
14	12 Januari 2012	ACC Keseluruhan		14

Malang, 13 Januari 2012

Mengetahui,
Ketua jurusan matematika,

Abdussakir, M.Pd
NIP: 19751006 200312 1 001