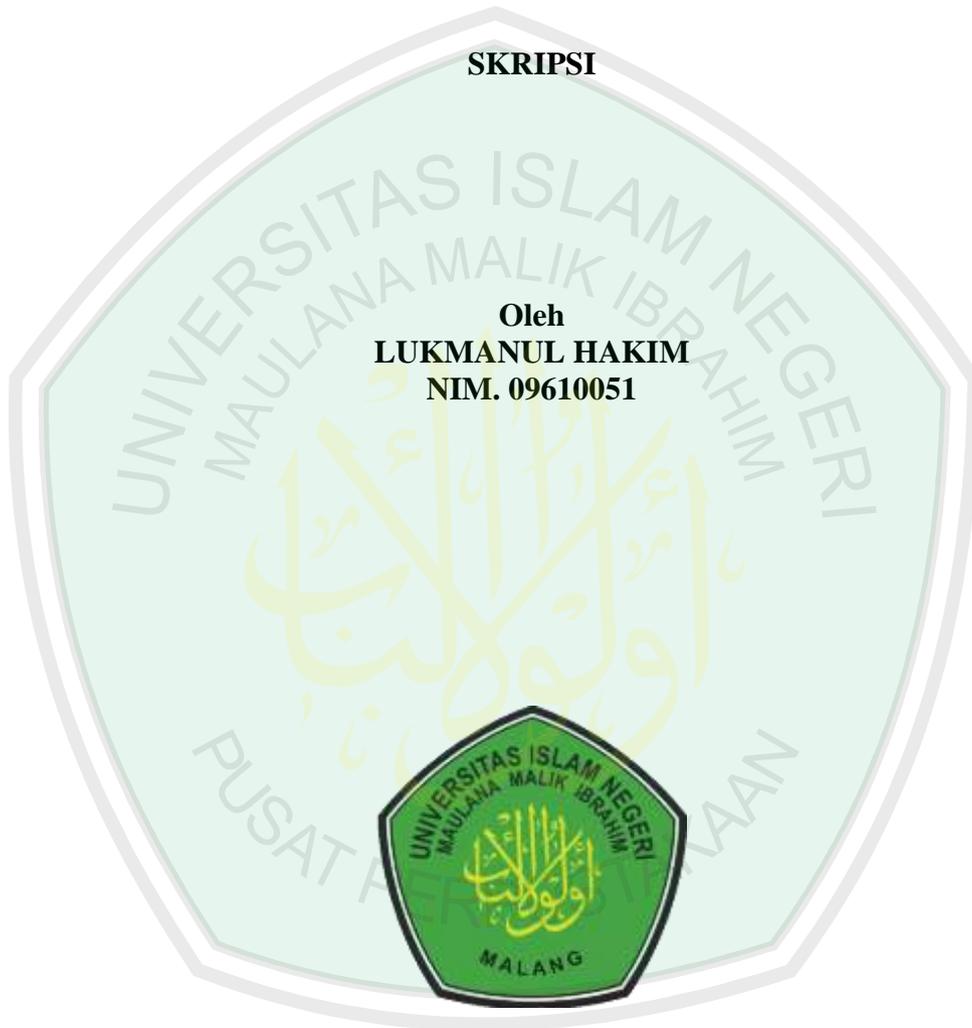


**ANALISIS REGRESI PADA DATA *OUTLIER* DENGAN METODE
MM-ESTIMASI**

SKRIPSI

Oleh
LUKMANUL HAKIM
NIM. 09610051



**JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2016**

**ANALISIS REGRESI PADA DATA *OUTLIER* DENGAN METODE
MM-ESTIMASI**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Sains (S.Si)**

**Oleh
Lukmanul Hakim
NIM. 09610051**

**JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2016**

**ANALISIS REGRESI PADA DATA *OUTLIER* DENGAN METODE
MM-ESTIMASI**

SKRIPSI

Oleh
Lukmanul Hakim
NIM. 09610051

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal 30 Juli 2016

Pembimbing I,

Pembimbing II,

Dr. Sri Harini, M.Si
NIP. 1971014 2001122 002

Evawati Alisah, M.Pd
NIP. 19720604 199903 2 001

Mengetahui,
Ketua Jurusan Matematika

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

PERNYATAAN KEASLIAN TULISAN

Saya yang bertandatangan di bawah ini:

Nama : Lukmanul Hakim

NIM : 09610051

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Analisis Regresi Pada Data *Outlier* Dengan Metode MM-Estimasi.

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

Malang, 30 Juni 2016

Yang membuat pernyataan,

Lukmanul Hakim

NIM. 09610051

MOTO

“Selalu berusaha melakukan yang terbaik meski manusia mempunyai kelebihan ataupun kekurangan, takdir bisa berubah jika kita berusaha”



PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

Ibu Munawaroh, bapak Bonasir (Alm.) dan adik tersayang Wardatul
Jannah, keluarga dari ibu penulis.

Bapak Moh. Sholeh, ibu Murniati dan adik Moh. Hadi Rizqi serta keluarga dari
bapak penulis.

Inspirator penulis yaitu Abah Romo Kyai Haji Prof. Abdul Ghofur pemangku
Pondok Pesantren Sunan Drajat Lamongan.



BAB IKATA PENGANTAR

Assalamu'alaikum Warohmatullahi Wabarokatuh.

Puji syukur kepada Allah Swt. Berkat rahmat dan izin-Nya penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang matematika di Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Dalam menyelesaikan skripsi ini, penulis banyak mendapat bimbingan dan arahan dari berbagai pihak. Untuk itu ucapan terimakasih yang sebesar-besarnya dan penghargaan setinggi-tingginya penulis sampaikan terutama kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. drh. Bayyinatul Muchtaromah, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Abdussakir, M.Pd, selaku ketua Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Sri Harini, M.Si., selaku dosen pembimbing I yang dengan sabar telah meluangkan waktunya demi membimbing, mengarahkan, menasihati serta memberim otivasi dalam penyelesaian skripsi ini.
5. Evawati Alisah, M.Pd., selaku dosen pembimbingII yang telah membimbing dan berbagi ilmu kepada penulis sampai menyelesaikan skripsi ini. Segenap sivitas akademika Jurusan Matematika Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang terutama seluruh dosen terimakasih atas ilmu dan bimbingan yang telah diberikan pada penulis.

6. Ibu penulis yaitu ibu Munawaroh, adik penulis yaitu Wardatul Jannah dan bapak Bonasir (Alm.) yang tidak pernah henti-hentinya memberikan kasih sayang, do'a, serta motivasi kepada penulis.
7. Bapak Moh. Sholeh dan ibu Murni serta adik Rizki.
8. Guru-guru Madrasah Ibtida'iyah, Madrasah Tsanawiyah, TPA, TPQ, MA. Ma'arif 07 Sunan Drajad yang telah membimbing penulis dari yang bukan siapa-siapa dan tidak mengerti apa-apa sampai bisa seperti ini. Sungguh jasa-jasa Beliau tidak ternilai harganya.
9. Semua mahasiswa Universitas Islam Negeri Maulana Malik Ibrahim Malang khususnya Jurusan Matematika terutama angkatan 2009 serta semua pihak yang tidak dapat penulis sebutkan satu persatu, atas keikhlasan bantuan moril maupun materiil, penulis ucapkan terimakasih.

Semoga skripsi ini bermanfaat bagi semua pihak dan menambah wawasan keilmuan khususnya di bidang matematika statistika. Aamiin ya Robbal 'Aalamiin.

Wassalamu 'alaikum Warohmatullahi Wabarokatuh.

Malang, Juni 2016

Penulis

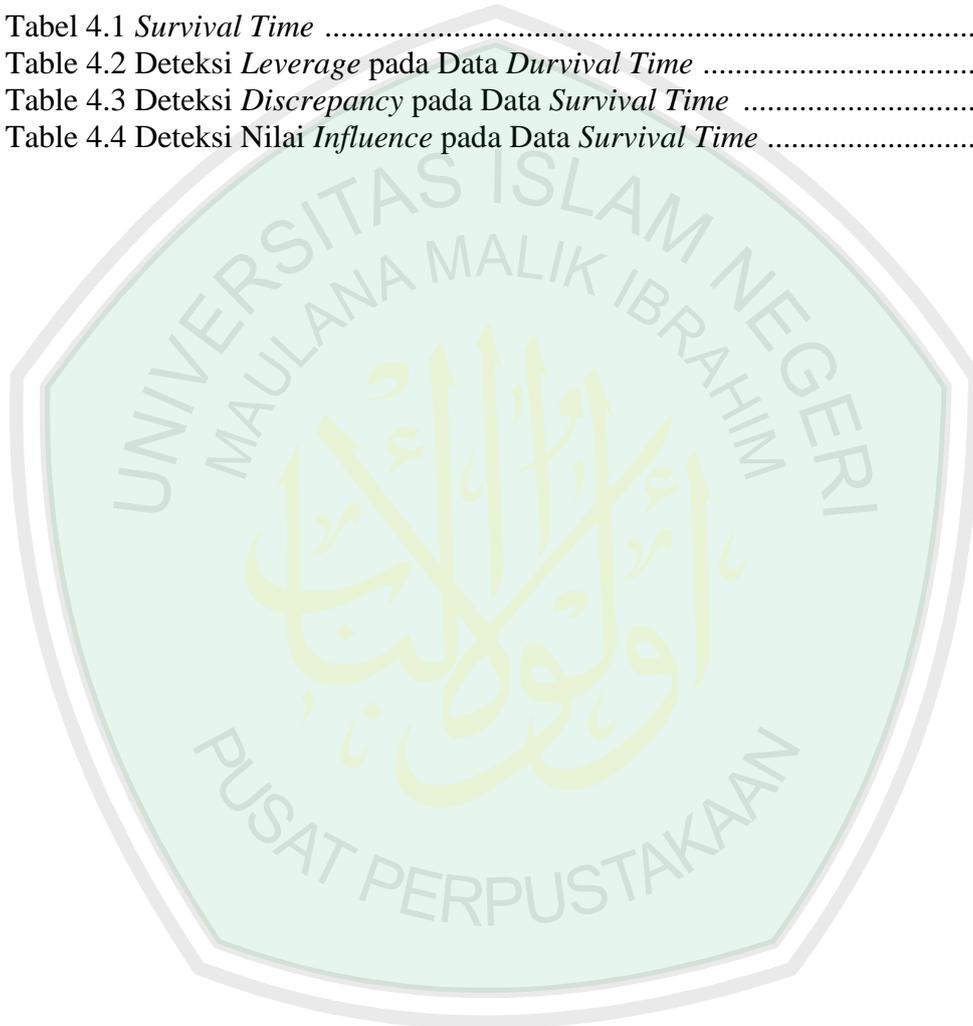
DAFTAR ISI

HALAMAN JUDUL	
HALAMAN PENGAJUAN	
HALAMAN PERSETUJUAN	
HALAMAN PENGESAHAN	
HALAMAN PERNYATAAN KEASLIAN TULISAN	
HALAMAN MOTO	
HALAMAN PERSEMBAHAN	
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR SIMBOL	xiii
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT	xvii
ملخص	xviii
BAB I PENDAHULUAN	
1.1	Lat
ar Belakang	1
1.2	Ru
musan Masalah	5
1.3	Tuj
uan Penelitian	5
1.4	Bat
asan Masalah.....	5
1.5	Ma
nfaat Penelitian	6
1.6	Sist
ematika Penulisan	7
BAB II KAJIAN PUSTAKA	
2.1	Ana
lisis Regresi	8

2.2	Model Persamaan Regresi	Mo
	del Persamaan Regresi	9
2.3	Outlier	Out
	lier	10
2.3.1	Nilai <i>Leverage</i>	12
2.3.2	Nilai <i>Discrepancy</i>	14
2.3.3	Nilai <i>Influence</i>	17
	2.3.3.1 Pengukuran nilai <i>influence</i> secara global	17
	2.3.3.2 Pengukuran nilai <i>influence</i> secara khusus	20
2.4	Metode Robust	Met
	ode Robust	21
2.4.1	MM-Estimasi	22
2.5	Kajian Al-Quran dan Al-Hadits Tentang Regresi dan <i>Outlier</i>	Kaji
	an Al-Quran dan Al-Hadits Tentang Regresi dan <i>Outlier</i>	24
2.5.1	Ayat Al-Quran Tentang Analisis Regresi	24
2.5.2	Ayat Al-Quran Tentang <i>Outlier</i>	26
BAB III METODE PENELITIAN		
3.1	Kerangka Penelitian	Ker
	angka Penelitian	29
3.2	Metode Pengumpulan Data	Met
	ode Pengumpulan Data	29
3.3	Analisis Data	Ana
	lisis Data	31
BAB IV PEMBAHASAN		
4.1	Deteksi <i>Outlier</i>	33
	4.1.1 <i>Leverage</i>	34
	4.1.2 <i>Discrepancy</i>	35
	4.1.3 <i>Influence</i>	36
4.2	MM-Estimasi	37
4.3	Kajian <i>Outlier</i> dalam Al-Quran	39
BAB V PENUTUP		
5.1	Kesimpulan	41
5.2	Saran	42
DAFTAR PUSTAKA		43
LAMPIRAN		44
RIWAYAT		HIDUP

DAFTAR TABEL

Tabel 4.1 <i>Survival Time</i>	30
Table 4.2 Deteksi <i>Leverage</i> pada Data <i>Durvival Time</i>	34
Table 4.3 Deteksi <i>Discrepancy</i> pada Data <i>Survival Time</i>	35
Table 4.4 Deteksi Nilai <i>Influence</i> pada Data <i>Survival Time</i>	37



DAFTAR SIMBOL

x	:	variabel bebas
y	:	variabel terikat
β_0	:	intercept pada sumbu y, titik potong dengan sumbu y
β_1	:	kemiringan dari garis regresi
ε	:	<i>error</i>
$y_{(n \times 1)}$:	vektor variabel terikat
$X_{(n \times (k+1))}$:	matriks variabel bebas
$\beta_{((k+1) \times 1)}$:	vektor koefisien parameter regresi
$W_1_{(n \times n)}$:	matriks pembobot <i>spatial</i>
$W_2_{(n \times n)}$:	matriks bobot <i>spatial error</i>
ρ	:	parameter koefisien <i>spatial lag</i> variabel dependen
λ	:	parameter koefisien <i>spatial lag error</i>
$u_{(n \times 1)}$:	vektor <i>error</i> yang diasumsikan mengandung autokorelasi
$\varepsilon_{(n \times 1)}$:	vektor <i>error</i> yang diasumsikan tidak mengalami autokorelasi, yang berdistribusi normal dengan <i>mean</i> nol dan <i>varians</i> $I\sigma^2$
$I_{(n \times n)}$:	matriks identitas
φ	:	fungsi objektif
ψ	:	fungsi <i>influence</i> (pengaruh)
ω	:	matriks pembobot
h_{ii}	:	<i>leverage</i> kasus ke-i

- n : banyaknya data
- X_i : nilai untuk kasus ke-i
- M_x : *mean* dari X
- Σx^2 : jumlah kuadrat n kasus dari simpangan X_i dari *meannya*



DAFTAR LAMPIRAN

Data <i>Survival Time</i>	30
Data Deteksi <i>Leverage</i> pada Data <i>Durvival Time</i>	34
Data Deteksi <i>Discrepancy</i> pada Data <i>Survival Time</i>	35
Data Deteksi Nilai <i>Influence</i> pada Data <i>Survival Time</i>	37



ABSTRAK

Hakim, Lukmanul. 2016. **Analisis Regresi pada Data *Outlier* dengan Metode MM-Estimasi**. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Sri Harini, M.Si, (II) Evawati Alisah, M.Pd.

Kata Kunci: Regresi, *Outlier*, *Robust* MM-Estimasi.

Model regresi digunakan untuk mempelajari hubungan antara sebuah variabel terikat (y) dan variabel bebas (x). Metode estimasi yang paling sering digunakan untuk menganalisis regresi adalah metode kuadrat terkecil. Metode kuadrat terkecil untuk model regresi linier dikenal sangat sensitif terhadap *outlier*. Salah satu alternatif untuk memperbaiki kelemahan metode kuadrat terkecil adalah menggunakan estimasi yang bersifat *robust* yang mampu bertahan terhadap kehadiran *outlier* pada data pengamatan.

Tugas akhir ini akan membahas salah satu metode regresi *robust* yaitu metode MM-estimasi (*Method of Moments*) yang bertujuan mengestimasi parameter regresi ketika data terkontaminasi *outlier*. Estimasi ini merupakan gabungan metode estimasi yang mempunyai nilai *breakdown* yang besar dan sifat *efisiensi* yang tinggi.

Tujuan penulisan tugas akhir ini adalah mendeteksi *outlier* dengan menggunakan nilai *leverage*, nilai *discrepancy* dan nilai *influence*. Setelah itu data dianalisis dengan menggunakan metode regresi *robust* MM-Estimasi. Data yang digunakan dalam penulisan ini adalah data sekunder yang diperoleh dari riset H. N. Cahya (2010) tentang *survival time*. Berdasarkan hasil analisis dapat disimpulkan bahwa penerapan metode MM-Estimasi pada *survival time* menghasilkan model:

$$\hat{y} = -353.6101 + 25.8923x_1 + 2.4334x_2 + 2.7282x_3 + 0.4754x_4$$

dan *R-square* yang dihasilkan adalah 94,1% serta menghasilkan estimasi *residual* sebesar 20,96. Hal tersebut menyatakan bahwa keakuratan metode MM-Estimasi tidak terpengaruh dengan adanya *outlier*, sehingga dapat diputuskan bahwa metode MM-Estimasi sangat cocok digunakan apabila dalam suatu data penelitian terdapat *outlier*.

ABSTRACT

Hakim, Lukmanul. 2016. **Regression Analyzes on Data of Outlier with MM-Estimation Method**. Thesis. Department of Mathematics, Faculty of Science and Technology, Islamic State University of Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Sri Harini, M.Si, (II) Evawati Aisah, M.Pd.

Keywords: Regression, Outlier, MM-Robust Regression Estimates.

Regression models were used to study the relationship between a dependent variable (y) and the independent variable (x). The estimation method is most often used to analyze the regression is the least squares method. The least squares method for the linear regression model known to be highly sensitive to outliers. One alternative to improve weaknesses in the least squares method is to use estimates that are robust which is able to withstand the presence of outliers in the data observations.

This final project will discuss a robust regression method is the method of MM-the estimates (Method of Moment) aimed at estimating the regression parameters when data have contamination outlier. This estimate is a combination of estimation methods that have great value and nature of the breakdown of high efficiency.

The purpose of writing this final task is to detect outliers by using leverage, the value of the discrepancies and values influence. After the data is analyzed using a robust regression method MM-estimation. The data used in this paper is secondary data obtained from research H. N. Cahya (2010) on survival time. Based on the analysis it can be concluded that the application of MM-estimation methods on survival time in a model:

$$\hat{y} = -353.6101 + 25.8923x_1 + 2.4334x_2 + 2.7282x_3 + 0.4754x_4$$

and R-square produced was 94.1% and generate an estimated residual of 20.96. It states that the accuracy of MM-estimation method is not affected by the presence of outliers, so it can be decided that MM-estimation method is suitable for use when in a research data are outliers.

ملخص

لقمان الحكيم. 2016. تحليل الانحدار في البيانات أوتلير مع أسلوب- تقدير م م (طريقة لحظة). بحث جامعي شعبة الرياضيات، كلية العلوم والتكنولوجيا، جامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: الدكتورة سري هاريني، الماجستير و إيفاواتي عالسة، الماجستير

كلمات الرئيسية: الانحدار، أوتلير، تقدير م م
واستخدمت نماذج الانحدار لدراسة العلاقة بين المتغير التابع (Y) والمتغير المستقل (X). تستخدم أسلوب التقدير لتحليل الانحدار هو طريقة المربعات الاصغر. طريقة أقل المربعات لنموذج الانحدار الخطي المعروف أن تكون حساسة للغاية لالتقييم أوتلير. بديل واحد لتحسين نقاط الضعف في طريقة المربعات الصغرى هو استخدام التقديرات التي هي قوية والتي هي قادرة على تحمل وجود القيم أوتلير في هذه الملاحظات البيانات.

هذا المشروع النهائي ومناقشة طريقة الانحدار القوي هو طريقة التقدير (طريقة لحظة)، والتي تهدف إلى تقدير معالم الانحدار عندما ملوثة البيانات مع القيم أوتلير. هذا التقدير هو مزيج من أساليب تقدير أن لها قيمة كبيرة وطبيعة اختيار كفاءة عالية.

والغرض من كتابة هذه المهمة الأخيرة هي للكشف عزلاء باستخدام القيم المتطرفة ، وقيمة تأثير والقيم التناقضات. بعد تحليل البيانات باستخدام قوي طريقة الانحدار تقدير طريقة لحظة . البيانات المستخدمة في هذه الورقة هو البيانات الثانوية التي تم الحصول عليها من البحث ه ن جهيا (2010) عن الوقت المحدد استنادا إلى تحليل يمكن الاستنتاج أن تطبيق أساليب تقدير طريقة لحظة في الوقت المحدد البقاء على قيد الحياة في نموذج:

$$\hat{y} = -353.6101 + 25.8923x_1 + 2.4334x_2 + 2.7282x_3 + 0.4754x_4$$

وكان R -square تنتج 94.1% وتوليد المتبقية المقدرة 20.96. وهو ينص على أن دقة طريقة تقدير طريقة لحظة لا يتأثر وجود القيم أوتلير ، لذلك يمكن أن يتقرر أن طريقة لحظة تقدير مناسب للاستخدام في حين بحث البيانات هي القيم أوتلير

BAB II

PENDAHULUAN

1.1 Latar Belakang

Statistik adalah cabang matematika yang berkaitan dengan pengumpulan data, pengolahan data, penyajian data, analisis data dan penarikan kesimpulan. Suatu kegiatan utama statistik adalah pengumpulan data. Dalam hal pengumpulan data yaitu mencatat atau pembukuan data, al-Quran juga membicarakannya dalam surat al-Kahfi ayat 49.

وَوُضِعَ الْكِتَابُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيْلَتَنَا مَا لِ هَذَا
الْكِتَابِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّا أَحْصَاهَا^ج وَوَجَدُوا مَا عَمِلُوا حَاضِرًا^ط
وَلَا يَظْلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

“Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang yang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya dan mereka berkata : “Aduhai celaka kami, kitab apakah ini yang tidak meninggalkan yang kecil dan tidak (pula) yang besar, melainkan ia mencatat semuanya” dan mereka dapati apa yang telah mereka kerjakan ada (tertulis). Dan Tuhanmu tidak menganiaya seorang juapun”(QS. al-Kahfi/18:49).

Dari ayat di atas menjelaskan keterkaitan antara isi kandungan surat al-Kahfi ayat 49 dengan matematika, yaitu pada khususnya statistik. Pada penggalan ayat terdapat kata alkitab yang berarti mencatat, sama halnya pada statistik langkah awalnya adalah mencatat terlebih dahulu data yang dibutuhkan, setelah mencatatnya barulah data tersebut dapat diolah. Pada ayat di atas juga terdapat kata ‘amilu yang berarti data. Jadi dalam statistik terdapat kegiatan mencatat dan mengumpulkan data setelah itu data tersebut dapat diolah sehingga diperoleh kesimpulan.

Model matematika dalam statistika merupakan penyederhanaan dari realitas atau permasalahan yang diteliti oleh statistikawan. Oleh karena itu, diperlukan asumsi-asumsi agar model tersebut dapat menggambarkan permasalahannya. Selain itu, asumsi diperlukan agar dapat merumuskan apa yang statistikawan ketahui atau terka mengenai penganalisisan data atau masalah pemodelan statistik yang dihadapinya dan pada saat yang bersamaan asumsi diperlukan agar model yang dihasilkan dapat memudahkan dalam sudut pandang teoritik dan komputasinya. Salah satu asumsi yang paling banyak ditemukan adalah asumsi kenormalan, yang telah ada selama dua abad, asumsi kenormalan menjadi kerangka berpikir dalam suatu metode statistik inferensi, yaitu : regresi, analisis variansi, analisis multivariate, model runtun waktu dan lain-lain.

Sering kali dalam praktenya asumsi kenormalan terpenuhi secara aproksimasi pada sebagian besar data observasi. Bahkan beberapa observasi berbeda pola atau bahkan tidak berpola mengikuti distribusi normal. Hal ini dikarenakan observasi yang tidak normal, observasi yang terpisah dari observasi-observasi lainnya yang dikenal dengan data *outlier*. Dengan data *outlier* asumsi kenormalan sering kali tidak terpenuhi, walaupun jumlah data cukup besar. Sehingga statistikawa kemungkinan melakukan kesalaha dalam memodelkan suatu fenomena dengan adanya kehadiran data *outlier*.

Ordinary Least Square adalah metode yang tepat untuk menyelesaikan model regresi dengan syarat semua asumsi harus terpenuhi, akan tetapi dalam kenyataannya sering kita temui ada sebagian asumsi dalam regresi yang tidak terpenuhi dikarenakan terdapatnya data *outlier*. Asumsi kenormalan seringkali tidak terpenuhi karena adanya pengamatan *outlier* yang memberikan pengaruh

besar terhadap estimasi parameter model. Jika asumsi kenormalan terpenuhi, maka metode *Ordinary Least Square* (OLS) dapat menduga dengan baik. Namun jika tidak terpenuhi, maka estimasi OLS tidak dapat digunakan (Montgomery, 1992).

Outlier itu sendiri adalah sebuah data pencilan dari kebanyakan data atau bisa diartikan juga bahwa *outlier* adalah sebuah data yang tidak mengikuti ritme sebuah alur dari kebanyakan data yang ada. Seperti yang telah didefinisikan oleh Barnett dan Lewis (1994), *outlier* adalah data yang muncul tidak konsisten dengan sisa data.

Dalam mengatasi data *outlier* harus dilihat dari sumber munculnya data yang menjadi *outlier* tersebut. *Outlier* mungkin ada karena adanya data terkontaminasi, yaitu adanya kesalahan pada saat melakukan pengambilan sampel pada populasi. *Outlier* yang disebabkan oleh data terkontaminasi dapat dihapuskan dari data penelitian atau jika memungkinkan dilakukan sampling ulang. Akan tetapi, jika setelah melakukan beberapa kali sampling ulang namun data *outlier* tetap muncul, maka data tersebut jangan dihapuskan dari data penelitian, karena analisis data yang dihasilkan akan tidak mencerminkan populasi yang diteliti.

Metode *Ordinary Least Square* tidak tepat untuk menyelesaikan data yang terdapat *outlier*, maka diperlukan metode lain supaya analisis data dengan hadirnya *outlier* tetap tahan terhadap asumsi yang diterapkan pada penganalisisan datanya. Metode tersebut dikenal dengan Metode *Robust*.

Model matematika dalam statistika yang banyak ditemukan dalam berbagai bidang adalah model regresi. Model regresi merupakan suatu model

yang menggambarkan hubungan dari variabel dependen dengan variabel-variabel independen. Dengan adanya hubungan tersebut diharapkan variabel independen dapat memprediksi nilai-nilai variabel dependen. Perlu diketahui bahwa data itu mempunyai pola bentuk. Berdasarkan pola bentuk kelinearan data, model regresi dapat diklasifikasikan menjadi dua macam yaitu linear dan non-linear. Jika pola data linear, maka digunakan pemodelan linear.

Model regresi mengasumsikan bahwa error dari model tersebut harus berdistribusi normal, bervariasi konstan dan saling independen antar observasi. Dengan adanya *outlier* pada data regresi mengakibatkan model regresi tidak memenuhi asumsi-asumsinya dan model regresi tidak cocok terhadap data yang akan dimodelkan karena nilai koefisien dari model regresi tersebut sangat dipengaruhi oleh adanya *outlier*. Oleh karena itu, model yang dihasilkan tidak dapat digunakan untuk memprediksi. Sehingga, *outlier* pada regresi harus diatasi. Salah satu metode guna mengatasi data *outlier* pada regresi adalah metode *robust*. Ada banyak macam metode *robust* diantaranya adalah *Least Trimmed Square (LTS) Estimate*, *M Estimate*, *MM Estimate*, *S Estimate* (Seber, 2007).

Salah satu metode *robust* yang mempunyai nilai breakdown point tinggi adalah MM-Estimasi yaitu hampir 50%. Dalam MM-Estimasi dibutuhkan iterasi awal (initial) dan iterasi akhir (final). Metode MM-Estimasi inilah yang akan dipakai oleh peneliti untuk menganalisis sekaligus mengatasi *outlier* pada model regresi.

Disamping penanganan *outlier* pada regresi, yang lebih penting adalah pengidentifikasian data yang menjadi *outlier*. Metode pengidentifikasian yang digunakan peneliti adalah dengan melihat *leverage*, nilai *discrepancy* dan nilai

influence-nya. *Leverage* hanya menggambarkan kasus yang terjadi pada variabel independen. Untuk tiap kasus, *leverage* menginformasikan seberapa jauh kasus tersebut dari nilai *mean* himpunan data variabel independen. Sedangkan *discrepancy* merupakan jarak antara nilai prediksi dengan nilai observasi dari variabel dependen (Y), yaitu $Y_i - \hat{Y}_i$ yang merupakan nilai dari residual e_i . Pada dasarnya nilai yang menjadi *outlier* menyebabkan nilai residual menjadi besar dan tidak jatuh pada garis regresi. Sedangkan nilai *influence* merupakan kombinasi dari ukuran *leverage* dan *discrepancy* yang menginformasikan mengenai bagaimana perubahan dari persamaan regresi jika kasus ke- i dihilangkan dari himpunan data (Cohen, 2003).

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah bagaimana menganalisis *outlier* dalam model regresi dengan metode MM-Estimasi.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah untuk dapat menganalisis *outlier* dalam model regresi dengan metode MM-Estimasi.

1.4 Batasan Masalah

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

1. asumsi regresi yang dipakai hanya asumsi kenormalan
2. untuk pengidentifikasian *outlier* dengan nilai *leverage* menggunakan metode *h-matrik*,
3. untuk pengidentifikasian *outlier* dengan nilai *discrepancy* menggunakan metode *externally studentized residual*,
4. untuk nilai *influence* diidentifikasi dengan metode *DFFIT'S* dan *Cook's*,
5. metode *robust* yang digunakan dalam penelitian ini adalah MM-Estimasi untuk menganalisis sekaligus mengatasi *outlier* pada model regresi.

1.5 Manfaat Penelitian

Skripsi ini diharapkan bermanfaat bagi berbagai kalangan, antara lain :

1. Bagi Penulis

Dapat mengaplikasikan ilmu yang telah diperoleh selama kuliah dan menambah ilmu pengetahuan dalam hal cara menganalisis sekaligus mengatasi *outlier* dalam model regresi dengan metode MM-Estimasi.

2. Bagi Pembaca

Dapat dijadikan sebagai tambahan referensi bagi mahasiswa matematika dalam memahami khususnya ilmu statistika dan aplikasinya dalam kehidupan.

3. Bagi Instansi

Sebagai tambahan bahan kepustakaan yang dapat dijadikan sebagai sarana pengembangan wawasan keilmuan khususnya di Jurusan Matematika mengenai ilmu statistika.

1.6 Sistematika Penulisan

Untuk mempermudah memahami skripsi ini secara keseluruhan, maka penulis menggunakan sistematika penulisan yang terdiri dari empat bab dan masing-masing akan dijelaskan sebagai berikut :

Bab I Pendahuluan

Pada bab ini diuraikan tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

Bab II Kajian Pustaka

Pada bab ini disajikan tentang konsep-konsep atau teori-teori yang mendukung bagian pembahasan.

Bab III Metode Penelitian

Pada bab ini akan dibahas tentang metode-metode penelitian untuk menjelaskan secara garis besar bagaimana langkah-langkah penelitian ini dilakukan.

Bab IV Pembahasan

Pada bab ini dijelaskan tentang pengaplikasian metode MM-Estimasi terhadap studi kasus pada persamaan model regresi yang terinfeksi *outlier* baik pada variabel dependen dan independen atau bahkan keduanya.

Bab V Penutup

Pada bab ini dipaparkan tentang kesimpulan dari hasil pembahasan setelah itu diberikan saran oleh penulis untuk pembaca sekalian.

BAB III

KAJIAN PUSTAKA

2.1 Analisis Regresi

Istilah regresi pertama kali diperkenalkan oleh Francis Galton dalam artikelnya "*family likenes 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 antar generasi. Hal ini dijelaskan Galton pada fakta yang memperlihatkan adanya kecenderungan mundurnya 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 dimaksud oleh Galton. Secara luas sekarang analisis regresi diartikan sebagai suatu analisis tentang ketergantungan suatu variabel kepada variabel lain dalam rangka membuat suatu estimasi atau prediksi dan rata-rata nilai variabel tergantung dengan diketahuinya nilai variabel bebas (Alghifari, 1997).

Secara umum ada dua macam hubungan antara dua variabel atau lebih, yaitu bentuk hubungan dan keeratan hubungan. Untuk mengetahui bentuk hubungan digunakan analisis regresi, sedangkan untuk keeratan hubungan dapat diketahui dengan analisis korelasi. Analisis regresi dipergunakan untuk menelaah hubungan antara dua variabel atau lebih, terutama untuk menelusuri pola hubungan yang modelnya belum diketahui dengan sempurna atau untuk

mengetahui bagaimana variasi dari beberapa variabel bebas mempengaruhi variabel terikat dalam suatu fenomena yang kompleks. Jika $X_1, X_2 \dots X_i$ adalah variabel bebas dan Y adalah variabel terikat, maka terdapat hubungan fungsional antara X dan Y , dimana variabel dari X akan diiringi pula oleh variabel dari Y .

Analisis regresi adalah teknik analisis yang mencoba menjelaskan bentuk hubungan antara peubah-peubah yang mendukung sebab akibat. Proses analisisnya didasarkan atas distribusi probabilitas bersama peubah-peubahnya. Bila hubungan ini dapat dinyatakan dalam persamaan matematika, maka dapat dimanfaatkan untuk keperluan-keperluan lain misalnya peramalan. Tujuan utama dari analisis regresi adalah mendapatkan dugaan (ramalan) dari suatu variabel dengan menggunakan variabel lain yang diketahui.

Untuk menguji model analisis regresi terdapat empat langkah antara lain (Wibisono, 2005):

1. Menentukan estimasi parameter dari model regresi,
2. Menguji normalitas data,
3. Menguji asumsi homoskedastisitas,
4. Menguji asumsi multikolinieritas.

2.2 Model Persamaan Regresi

Regresi merupakan suatu alat ukur untuk mengukur ada atau tidak adanya hubungan antara variabel bebas (X) dan variabel terikat (Y). Istilah regresi yang berarti ramalan atau taksiran pertama kali diperkenalkan oleh Sir Francis Galton

(1877). Dengan mengetahui adanya hubungan antara variabel tersebut dapat dilakukan pendugaan suatu variabel berdasarkan variabel lain melalui persamaan yang dihubungkan tersebut (Alghifari, 1997).

Model regresi linier secara umum dapat dinyatakan dengan:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.1)$$

dimana:

y = variabel terikat

x = variabel bebas

β_0 = *intercept* pada sumbu y , titik potong dengan sumbu y

β_1 = kemiringan dari garis regresi

ε = *error*.

2.3 Outlier

Outlier adalah pengamatan yang berada jauh (ekstrim) dari pengamatan-pengamatan lainnya. *Outlier* juga dapat diartikan data yang tidak mengikuti pola umum pada model atau yang keluar pada model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 1995). Oleh karena itu apabila dalam suatu data pengamatan terdapat *outlier*, maka untuk mencari solusinya tidak diperkenankan menggunakan metode *Least Square Estimate* (LS) karena metode LS didasarkan pada asumsi bahwa error dari model yang dihasilkan harus berdistribusi normal.

Error yang merupakan *outlier* adalah yang nilai mutlaknya jauh lebih besar dari pada error lainnya dan bisa jadi terletak tiga atau empat kali simpangan baku atau lebih jauh lagi dari rata-rata errornya. *Outlier* merupakan suatu

keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya (Draper dan Smith 1998).

Sebagaimana yang telah dikemukakan Soemarti (2007) bahwa Ferguson mendefinisikan *outlier* sebagai suatu pengamatan yang menyimpang 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. *Outlier* merupakan nilai ekstrim dari suatu pengamatan. Seperti yang diketahui bahwa pada analisis regresi terdapat suatu variabel dependen dan independen yang digambarkan dalam scatterplot sebagai arah x dan y, oleh karena itu kemungkinan keberadaan *outlier* bisa terjadi pada arah x atau y dan bahkan bisa terjadi pada arah x dan y sekaligus.

Apabila data *outlier* terdapat pada arah x, maka akan memberikan pengaruh yang sangat besar pada estimator metode LS karena *outlier* pada arah x akan membalikkan garis ideal (LS). Oleh karena itu, *outlier* pada arah x disebut dengan titik *leverage*. Sedangkan apabila data *outlier* itu terletak pada arah y, maka akan memberikan nilai residual r yang sangat besar (negatif atau positif) sehingga tidak menunjukkan garis LS (Rousseuw, 1986).

Untuk mengetahui ada atau tidak adanya data *outlier* dapat dicari dengan menggunakan diagram pencar dari variabel independen maupun dependen atau dapat juga dicari dengan pemeriksaan visual dari data mentahnya, akan tetapi apabila dalam suatu kasus ketika terdapat lebih dari dua variabel independen,

maka akan sulit ditemukan data yang mengandung *outlier* menggunakan menggunakan pemeriksaan visual. Oleh karena itu, oleh karena itu dibutuhkan alat bantu dengan menggunakan uji statistik tertentu yang dikenal dengan regresi diagnostik yang dapat membantu dalam pendeteksian *outlier*.

Regresi diagnostik merupakan kasus statistik, artinya akan terdapat satu nilai dari setiap kasus diagnostik statistik dari setiap n-kasus dalam himpunan data. Semisal dalam suatu sampel dengan 200 kasus, maka akan menghasilkan 200 nilai dari setiap diagnostik statistiknya dan masing-masing nilai akan mempresentasikan setiap kasus dalam himpunan data tersebut. Fungsi regresi diagnostik adalah digunakan untuk memeriksa tiga karakteristik yang berpotensi merupakan data yang mengandung *outlier*. Yang pertama adalah *leverage*, yang mana akan mendiskripsikan seberapa tidak biasanya kasus atau data tersebut dalam ruang lingkup variabel independennya (x). Yang kedua adalah *discrepancy*, jarak antara nilai prediksi dan nilai observasi pada variabel hasil (y). Dan yang terakhir adalah nilai *influence*, yang mana secara konseptual *influence* mempresentasikan perkalian antara nilai *leverage* dan nilai *discrepancy*. Ketiga karakteristik ini harus diperiksa secara detail terlebih dahulu karena ketiga karakteristik tersebut akan menentukan letak dari setiap kasus data yang mengandung *outlier*. Artinya dengan ketiga karakteristik tersebut bisa mencari titik dimana data *outlier* itu berada, apakah dalam scatterplot x atau scatterplot y atau dalam keduanya (Cohen, 2003).

2.3.1 Nilai *Leverage*

Secara fungsi umumnya nilai *leverage* akan menggambarkan suatu kasus yang terletak dalam scatterplot x atau variabel independennya dalam suatu regresi.

Untuk setiap kasus *leverage* menginformasikan seberapa jauh kasus tersebut dari nilai *mean* himpunan data variabel independen. Seperti yang sudah dijelaskan sebelumnya bahwa dalam sebuah kasus terdapat sebuah data yang hanya mempunyai satu variabel independen dan ada pula data yang mengandung lebih dari dua variabel independennya. Jika hanya terdapat satu variabel independen, maka *leverage* dapat dipersamakan sebagai (Cohen, 2003):

$$\text{Leverage} = h_{ii} = \frac{1}{n} + \frac{(X_i - M_x)^2}{\Sigma x^2} \quad (2.2)$$

dengan : h_{ii} = *leverage* kasus ke-i

n = banyaknya data

X_i = nilai untuk kasus ke-i

M_x = *mean* dari X

Σx^2 = jumlah kuadrat n kasus dari simpangan X_i dari *meannya*

Jika kasus ke-i bernilai M_x bentuk persamaan di atas akan 0 dan h_{ii} akan memiliki nilai kemungkinan minimum $\frac{1}{n}$. Misalkan kasus ke-i skor pada X menjadi jauh dari M_x , maka akan menaikkan h_{ii} . Nilai maksimum dari h_{ii} adalah 1 dan nilai *mean* dari *leverage* untuk n kasus dalam suatu sampel adalah $M_{h_{ii}} = (k + 1)/n$, dengan k merupakan jumlah dari variabel independen dan n adalah banyaknya data. Perhitungan di atas dikhususkan untuk kasus dengan satu variabel independen. Sedangkan untuk kasus dengan lebih dari satu variabel independen, maka yang menjadi menarik adalah seberapa jauh nilai-nilai untuk k variabel untuk kasus ke-i, misalkan $X_{i1}, X_{i2}, X_{i3}, \dots, X_{ik}$ dari *centroid* variabel

independen, *centroid* merupakan *mean* dari data, $M_1, M_2, M_3, \dots, M_k$. perhitungan nilai h_{ii} untuk kasus ini dengan menggunakan persamaan (Weisberg, 2005).

$$H = X(X'X)^{-1}X' \quad (2.3)$$

Dengan H merupakan matrik $n \times n$ sedangkan X adalah matrik $n \times (k + 1)$.

Dengan n merupakan banyaknya data dan k adalah jumlah koefisien (β_k) variabel independen ditambah 1 sebagai nilai konstanta (β_0). Diagonal dari nilai H berisi nilai-nilai *leverage*. Jadi, untuk *leverage* kasus ke-i, h_{ii} , merupakan nilai dari baris ke-i dan kolom ke-i dari H.

Penentuan nilai yang memiliki *leverage* yang besar didasarkan pada nilai *cutoff*. Nilai h_{ii} yang melebihi nilai *cutoff* dideteksi sebagai *outlier*. Adapun nilai *cutoff* yang telah ditentukan adalah $2(k + 1)/n$ untuk data yang banyak $n > 15$, sedangkan untuk data yang sedikit $n \leq 15$ digunakan nilai *cutoff* $3(k + 1)/n$ (Cohen, 2003).

2.3.2 Nilai *Discrepancy*

Langkah kedua mendiagnostik statistik untuk data *outlier* adalah *discrepancy* antara nilai prediksi dan nilai observasi dari variabel dependen yaitu $Y_i - \hat{Y}_i$ yang mana merupakan nilai residual e_i . Pada dasarnya nilai yang menjadi *outlier* menyebabkan nilai residual menjadi besar dan tidak jatuh pada garis

regresi. Untuk mencari dari nilai *discrepancy* dilakukan dengan dua metode yaitu *internally studentized residuals* dan *externally studentized residuals*.

Yang dimaksud *internally studentized residual* adalah suatu metode yang menunjukkan satu dari dua hal yang menyangkut *residual* mentah (*raw*). Variansi *residual* untuk kasus ke-*i* diekspektasikan sebagai (Cohen, 2003).

$$\text{Variansi } (e_i) = MS_{\text{residual}}(1 - h_{ii}) \quad (2.4)$$

dengan MS_{residual} merupakan estimasi dari keseluruhan variansi dari *residual* sekitar garis regresi $(1 - R^2)(\Sigma y^2)/(n - k - 1)$. h_{ii} merupakan *leverage* dari kasus ke-*i*.

Standart deviasi dari *residual* kasus ke-*i* adalah

$$sd_{e_i} = \sqrt{MS_{\text{residual}}(1 - h_{ii})} \quad (2.5)$$

internally studentized residual merupakan rasio dari besaran nilai *residual* dari kasus ke-*i* dengan standart deviasi dari *residual* kasus ke-*i* yaitu (Cohen, 2003).

$$\text{internally studentized residuals } i = \frac{e_i}{sd_{e_i}} \quad (2.6)$$

Nilai dari *internally studentized residuals* berjarak antara 0 dan $\sqrt{n - k - 1}$ hal ini sangat tidak menguntungkan karena *internally studentized residuals* tidak mengikuti distribusi standart statistik karena persamaan (2.6) penyebut dan pembilangnya tidak saling bebas. Jadi *internally studentized residuals* tidak bisa diinterpretasikan menggunakan kurva normal atau t table. Oleh sebab itu metode tersebut jarang digunakan sehingga metode kedua yaitu *externally studentized residuals* lebih sering digunakan untuk menghitung nilai *discrepancy*.

externally studentized residuals merupakan metode yang kedua untuk perhitungan data yang mengandung *outlier* dalam lingkup *discrepancy*. *externally studentized residuals* menunjukkan permasalahan terhadap apa yang terjadi jika kasus yang dianggap *outlier* dihapuskan dari himpunan data pengamatan. Misalkan $\hat{Y}_{(i)}$

nilai prediksi kasus ke- i , akan tetapi pengamatan kasus ke- i dihapuskan dari himpunan data pengamatan. *Outlier* berkontribusi secara substansial terhadap estimasi variansi *residual* disekitar garis regresi, $MS_{residual}$. Sedangkan

$MS_{residual(i)}$ untuk variansi *residual* dengan kasus ke- i yang merupakan *outlier* dihapuskan dari himpunan data pengamatan. Misalkan d_i sebagai perbedaan antara

data yang asli dari pengamatan dengan nilai prediksi untuk kasus ke- i yang berasal dari himpunan data pengamatan dengan kasus ke- i yang dihapuskan yaitu $d_i = Y_i - \hat{Y}_i$. *externally studentized residuals* untuk kasus ke- i , t_i dihitung sebagai

berikut (Cohen, 2003):

$$t_i = \frac{d_i}{SE_{d_i}} \quad (2.7)$$

Untuk paralel dari persamaan (2.6) dan pembilang dari persamaan (2.7) merupakan *residual* untuk kasus ke- i dihapuskan dan penyebut merupakan standar *error* dengan kasus ke- i juga dihapuskan. *Residual* yang dihapuskan, d_i , dapat

dihitung dengan menggunakan *residual* awal, e_i , yaitu dengan persamaan sebagai

berikut:

$$d_i = \frac{e_i}{1-h_{ii}} \quad (2.8)$$

sedangkan untuk nilai standar *residual* dapat dihitung dengan persamaan berikut:

$$SE_{e_i} = \sqrt{\frac{MS_{residual(i)}}{1-h_{ii}}} \quad (2.9)$$

jika persamaan (2.8) dan (2.9) dimasukkan kedalam persamaan (2.7), maka t_i akan menjadi:

$$t_i = \frac{e_i}{\sqrt{MS_{residual(i)}(1-h_{ii})}} \quad (2.10)$$

Metode untuk penentuan nilai *outlier* berdasarkan nilai *externally studentized residuals* lebih sering digunakan dari pada metode *internally studentized residuals* karena jika model regresi cocok dengan data, maka *externally studentized residuals* akan mengikuti pola distribusi t dengan $df = n - k - 1$. Penentuan nilai *cutoff*-nya berdasarkan distribusi t , jika nilai t_i lebih besar dari nilai t_{tabel} dengan derajat kepercayaan α , maka data tersebut memiliki nilai *discrepancy* yang besar dan dikategorikan sebagai *outlier* (Cohen, 2003).

Perbedaan signifikan antara metode *internally studentized residuals* dengan metode *externally studentized residuals* adalah terletak pada interpretasi kurva normal atau t table. Bahwa *internally studentized residuals* tidak bisa diinterpretasikan menggunakan kurva normal atau t table karena suatu hal tertentu, sedangkan *externally studentized residuals* sebaliknya. Perbedaan tersebut yang menyebabkan metode *externally studentized residuals* lebih sering digunakan dibandingkan dengan metode *internally studentized residuals* dalam menentukan nilai *discrepancy*.

2.3.3 Nilai Influence

Ukuran nilai *Influence* merupakan kombinasi dari ukuran nilai *leverage* dan nilai *discrepancy* yang menginformasikan mengenai bagaimana perubahan dari persamaan regresi jika kasus ke-*i* dihilangkan dari himpunan data. Dua jenis metode yang biasa digunakan untuk pengukuran nilai *influence* ialah yang pertama ukuran ke-*influence*-an global (*DFFITs* dan *Cook'sD*) yang memberikan informasi bagaimana kasus ke-*i* mempengaruhi keseluruhan karakteristik dari persamaan regresi. Sedangkan yang kedua adalah ukuran ke-*influence*-an khusus (*DFBETAS*) yang menginformasikan mengenai bagaimana kasus ke-*i* mempengaruhi tiap-tiap koefisien regresi (Cohen, 2003).

2.3.3.1 Pengukuran nilai *influence* secara global

ada dua metode untuk melakukan pengukuran nilai *influence* yaitu *DFFITs* dan *Cook'sD*. Keduanya merupakan aspek yang membandingkan persamaan regresi yang mana ketika untuk kasus ke-*i* dihapuskan dan tidak dihapuskan dalam perhitungan himpunan data dalam penelitian.

Yang pertama untuk pengukuran nilai ke-*influence*-an adalah dengan metode *DFFITs*, yang mana secara teori didefinisikan sebagai berikut (Cohen, 2003):

$$DFFITs_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MS_{residual(i)} h_{ii}}} \quad (2.11)$$

dengan \hat{Y}_i merupakan nilai prediksi ketika kasus ke-*i* tidak dihapuskan atau dimasukkan kedalam himpunan data penelitian, $\hat{Y}_{i(i)}$ adalah nilai prediksi ketika kasus ke-*i* dihapuskan atau tidak dimasukkan kedalam himpunan data penelitian,

$MS_{residual(i)}$ merupakan nilai variansi dari *residual* ketika kasus ke- i dihapuskan dari himpunan data penelitian sedangkan h_{ii} adalah merupakan nilai *leverage*.

Menurut penjelasan di atas, bahwa ukuran nilai *influence* adalah merupakan perkalian antara nilai *leverage* dan nilai *discrepancy* sehingga *DFFITs* dapat ditulis seperti persamaan berikut (Cohen, 2003):

$$DFFITs_i = t_i \sqrt{\frac{h_{ii}}{\sqrt{1-h_{ii}}}} \quad (2.12)$$

dengan t_i adalah *externally studentized residuals* yang terdapat pada persamaan (2.13), sedangkan h_{ii} adalah nilai dari *leverage* yang terdapat pada persamaan (2.2 dan 2.3). Jika nilai dari t_i dan h_{ii} keduanya naik, maka besar nilai dari *DFFITs* juga akan naik. Hal ini menunjukkan kasus tersebut mempunyai pengaruh yang besar pada hasil analisis regresi. *DFFITs* = 0 ketika kasus ke- i persis terletak pada garis regresi, sehingga nilai \hat{Y}_i tidak mengalami perubahan ketika kasus ke- i tidak dimasukkan atau dihapuskan. Akan tetapi jika terletak pada *centroid* data sampel, maka akan memberikan beberapa pengaruh (*influence*) dikarenakan nilai minimum dari h_{ii} adalah $\frac{1}{n}$. Tanda untuk nilai *DFFITs* $\hat{Y}_i > \hat{Y}_{i(i)}$ adalah positif dan juga sebaliknya, jika tanda untuk nilai *DFFITs* $\hat{Y}_i < \hat{Y}_{i(i)}$, maka akan bernilai negatif.

Ukuran kedua untuk mengukur ke-*influece*-an secara global pada hasil model regresi dikarenakan kasus ke-*i* adalah dengan metode *Cook'sD*, yang didefinisikan sebagai berikut (Cohen, 2003):

$$Cook'sD_i = \frac{\sum(\hat{Y}_i - \hat{Y}_{i(i)})^2}{(k+1)MS_{residual}} \quad (2.13)$$

dengan \hat{Y}_i merupakan nilai prediksi ketika nilai kasus ke-*i* dimasukkan kedalam himpunan data penelitian, sedangkan $\hat{Y}_{i(i)}$ merupakan nilai prediksi ketika kasus ke-*i* dihilangkan dari himpunan data penelitian, *k* merupakan jumlah koefisien model regresi, dan $MS_{residual}$ merupakan nilai variansi dari *residual*. Secara garis besar metode *Cook'sD* membandingkan nilai prediksi dari *Y* dengan kasus ke-*i* dimasukkan dan dihapuskan dari data. Interpretasi untuk penyebut pada persamaan (2.16) di atas memberikan nilai standardisasi. Maksudnya tidak seperti metode *DFFITD*, *Cook'sD* akan selalu mempunyai nilai ≥ 0 , maka tidak bisa negatif.

DFFITS dan *Cook'sD* adalah merupakan dua metode ukuran yang saling berhubungan antara satu sama lainnya. Oleh karena itu, *DFFITS* dan *Cook'sD* mempunyai persamaan matematik sebagai berikut (Cohen, 2003):

$$Cook'sD_i = \frac{(DFFITS)_i^2 MS_{residual(I)}}{(k+1)MS_{residual}} \quad (2.14)$$

Kedua metode (*DFFITS* dan *Cook'sD*) dapat digunakan untuk memberikan informasi mengenai ke-*influence*-an dari kasus ke-*i* yang merupakan *outlier*. Penentuan untuk kasus ke-*i* sebagai *outlier* berdasarkan nilai *cutoff* masing-masing. Untuk *DFFITS* nilainya dengan mengabaikan tandanya yang besarnya > 1 untuk data ukuran kecil ($n \geq 15$) data ukuran sedang dideteksi

sebagai *outlier*, sedangkan untuk data berukuran besar nilai $DFFITs > 2\sqrt{\frac{(k+1)}{n}}$

merupakan data *outlier*. Untuk metode *Cook'sD* digunakan nilai *cutoff* 1.0 atau dengan nilai kritik dari distribusi F dengan $\alpha = 0,5$ sedangkan $df = (k + 1, n - k - 1)$. Jika nilai *Cook'sD* melebihi nilai kritik dari distribusi F ,

maka dideteksi sebagai *outlier* (Cohen, 2003).

2.3.3.2 Pengukuran nilai *influence* secara khusus

Pengukuran nilai *influence* secara khusus yaitu bisa disebut dengan metode $BF\beta_{TAS_{ij}}$ merupakan jenis kedua dari ke-*influence*-an statistik yang penting jika ingin memfokuskan pada koefisien regresi tertentu dalam persamaannya. $BF\beta_{TAS_{ij}}$ mempunyai fungsi yang sangat penting yaitu membandingkan koefisien-koefisien regresi ketika kasus ke- i yang mengandung *outlier* dimasukkan dalam himpunan data atau tidak dimasukkan dalam himpunan data penelitian.

Persamaan $BF\beta_{TAS_{ij}}$ untuk kasus ke- i dapat didefinisikan menjadi

(Cohen 2003):

$$BF\beta_{TAS_{ij}} = \frac{\beta_j - \beta_{j(i)}}{SE_{\beta_{j(i)}}} \quad (2.15)$$

Pada persamaan di atas, pembilang merupakan perbedaan dari koefisien dengan seluruh data yang dimasukkan (β_j), dengan koefisien jika kasus ke- i dihilangkan ($\beta_{j(i)}$). Sedangkan penyebut $SE_{\beta_{j(i)}}$ merupakan standart *error* dari

$\beta_{j(i)}$ setelah data ke- i dihilangkan. Pembandingan dengan $SE_{\beta_{j(i)}}$ memberikan nilai yang distandardisasi, fungsinya untuk menginterpretasi secara umum pengaruh dari kasus ke- i untuk semua koefisien regresi. Tiap kasus data akan memiliki $(k + 1) BFBETAS_{ij}$ yang berkorespondensi dengan tiap koefisien regresi dalam persamaannya termasuk *intercept* (β_0).

Penentuan nilai kasus yang memiliki ke-*influence*-an yang merupakan *outlier* berdasarkan $BFBETAS_{ij}$ adalah kasus yang memiliki $BFBETAS_{ij} > \pm 1$ untuk ukuran sampel yang kecil dan sedang, sedangkan penentuan dengan *cutoff* $BFBETAS_{ij} > \pm \frac{2}{\sqrt{n}}$ untuk ukuran sampel besar (Cohen, 2003).

2.4 Metode *Robust*

Regresi *robust* diperkenalkan Andrews (1972). Metode ini merupakan alat penting untuk menganalisis data yang dipengaruhi oleh *outlier* untuk menghasilkan model yang *robust* atau *resistant* terhadap *outlier*. Suatu estimasi yang *resistant* adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data. Metode *robust* ditujukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data *outlier* dan juga bersifat otomatis dalam menanggulangi data *outlier* (Aunuddin, 1989). Sifat efisiensi, titik kerusakan (*breakdown point*), dan titik *leverage* yang tinggi digunakan untuk menentukan kinerja teknik *robust* dalam arti teoritis.

Breakdown point adalah salah satu cara untuk mengukur ke-*robust*-an suatu estimator. Breakdown juga merupakan ukuran proporsi minimal dari banyaknya data yang terkontaminasi pencilan dibandingkan seluruh data pengamatan. Beberapa estimator regresi kemungkinan memiliki breakdown point paling sedikit $1/n$ atau $0/n$. Dengan kata lain, hanya dengan satu *outlier* akan menyebabkan persamaan regresi yang akan diberikan tidak berguna. Penduga lain kemungkinan memiliki breakdown point tertinggi $n/2$ atau 50%. Jika teknik estimasi *robust* memiliki 50% breakdown point, maka 50% dari data bisa berisi *outliers* dan koefisien akan tetap bisa digunakan. Titik *leverage* adalah pengamatan (x_e, y_p) yang setiap kali x_p terletak jauh dari sebagian besar x_i dalam sampel.

Chen (2002) menyebutkan beberapa metode estimasi parameter dalam regresi *robust* yaitu Least Trimmed Squares (LTS) Estimation, M Estimation, Yohai MM Estimation, S Estimation, Least Mean Square (Seber, 2007). Yang digunakan peneliti untuk melakukan penelitian adalah dengan menggunakan metode MM-Estimasi.

2.4.1 MM-Estimasi

Metode MM-Estimasi dikenalkan oleh Yohai pada tahun 1987 yang menghubungkan suatu *high breakdown point* 50% dengan efisien tinggi mencapai 95% (Wilcox, 2005).

MM-Estimasi dimulai dengan mencari S Estimasi yang sangat *robust* dan resisten yang meminimumkan suatu skala residual. Selanjutnya skala residual tetap konstan dan diakhiri dengan menetapkan parameter-parameter regresi

menggunakan M Estimasi. MM-Estimasi mempunyai *breakdown point* yang tinggi sama dengan S Estimasi yaitu sebesar 0,5 atau 50%, sehingga MM-Estimasi dapat menjelaskan bahwa banyaknya *outlier* hingga separuh data pengamatan tidak berpengaruh terhadap MM-estimasi. Berikut formula MM-Estimasi dapat dijelaskan seperti dibawah ini:

$$\hat{\beta}_{mm} = \min \sum \rho \left(\frac{e_i}{\hat{\sigma}} \right) = \min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=1}^n x_{ij} \beta_j}{\hat{\sigma}} \right) \quad (2.16)$$

S Estimasi sebagai permulaan dengan nilai high breakdown yang tinggi (50%) dan diakhiri dengan M Estimasi yang membuatnya mempunyai efisiensi yang tinggi. Pada umumnya digunakan fungsi Tukey Bisquare baik pada S Estimasi maupun M Estimasi (Wilcox, 2005).

Sebagaimana dalam kasus M Estimasi, MM-Estimasi juga menggunakan metode *Iteratively Reweighted Least Square* (IRLS) untuk mencari estimasi parameter regresi. Prosedur MM-Estimasi dapat diuraikan sebagai berikut:

1. Mengestimasi koefisien $\hat{\beta}_j^{(1)}$, sehingga diperoleh residual $e_i^{(1)}$ yang diambil dari regresi *robust* dengan *high breakdown point*.

2. Residual $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala residual M Estimasi, $\hat{\sigma}$ dan dihitung pula bobot awal $w_i^{(1)}$.

3. Residual $e_i^{(1)}$ dan skala residual $\hat{\sigma}$ dari langkah kedua digunakan dalam iterasi awal dengan metode WLS untuk menghitung koefisien regresi $\sum_{i=1}^n w_i^{(1)} \left(\frac{e_i^{(1)}}{\hat{\sigma}} \right) x_i = 0$ dimana $w_i^{(1)}$ menggunakan pembobot Huber atau Tukey Bisquare.

4. Menghitung pembobot baru $w_i^{(1)}$ menggunakan residual dari iterasi awal

WLS (langkah ketiga)

5. Langkah kedua, ketiga dan keempat diulang (reiterasi dengan skala residual tetap konstan) sampai $\sum_{i=1}^n |e_i^{(m)}|$ konvergen, yaitu selisih β_j^{m+1} dengan

β_j^m kurang dari 10^{-4} atau sampai mendekati 0 (m = banyaknya iterasi), maka

iterasi dihentikan (Yohai, 1987).

2.5 Kajian Al-Quran dan Al-Hadits Tentang Regresi dan *Outlier*

Statistik adalah cabang matematika yang berkaitan dengan pengumpulan data, pengolahan data, analisis data dan penarikan kesimpulan. Kegiatan utama dalam statistik adalah pengumpulan data, hal ini dibicarakan al-Quran dalam surat al-Kahfi ayat 49.

وَوُضِعَ الْكِتَابُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيْلَتَنَا مَا لِ هَذَا
 الْكِتَابِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّا أَحْصَاهَا وَوَجَدُوا مَا عَمِلُوا حَاضِرًا
 وَلَا يَظْلِمُ رَبُّكَ أَحَدًا

“Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang yang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya dan mereka berkata : “Aduhai celaka kami, kitab apakah ini yang tidak meninggalkan yang kecil dan tidak (pula) yang besar, melainkan ia mencatat semuanya” dan mereka dapati apa yang telah mereka kerjakan ada (tertulis). Dan Tuhanmu tidak menganiaya seorang juapun”(QS. al-Kahfi/18:49).

Dari ayat di atas menjelaskan keterkaitan antara isi kandungan surat al-Kahfi ayat 49 dengan matematika, yaitu pada khususnya statistik. Pada penggalan ayat terdapat kata *alkitabu* yang berarti mencatat, sama halnya pada statistik

langkah awalnya adalah mencatat terlebih dahulu data yang dibutuhkan, setelah mencatatnya barulah data tersebut dapat diolah. Pada ayat di atas juga terdapat kata ‘*amilu* yang berarti data. Jadi dalam statistik terdapat kegiatan mencatat dan mengumpulkan data setelah itu data tersebut dapat diolah sehingga diperoleh kesimpulan.

2.5.1 Ayat Al-Quran Tentang Analisis Regresi

Al-Quran merupakan kitab Allah yang didalamnya terkandung ilmu-ilmu Allah. Untuk mendapatkan ilmu tersebut perlu mengkaji al-Quran secara mendalam. al-Quran surat al-Baqarah ayat 2-3 dapat digunakan untuk analisis regresi dengan cara mempartisinya (membagi) dan hasil partisian ayat tersebut dimisalkan dengan sebuah variabel, ayat tersebut berbunyi:

ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ ﴿٢﴾ الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ
وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ ﴿٣﴾

“Kitab al-Qur’an ini tidak ada keraguan padanya : petunjuk bagi mereka yang bertaqwa. Yaitu mereka yang beriman kepada yang ghoib, mendirikan shalat dan menafkahkan sebagian rezeki yang kami anugerahkan kepada mereka”(QS. al-Baqarah/2:2-3).

Dalam ayat tersebut dijelaskan bahwa tidak ada keraguan didalam kitab suci al-Quran. Al-Quran ini juga merupakan petunjuk bagi mereka yang bertaqwa (dianggap *Y* variabel respon). Sedangkan kriteria taqwa itu adalah gabungan orang-orang yang mempunyai karakter ‘beriman kepada yang ghoib, yang mendirikan shalat dan menafkahkan sebagian rezeki yang dianugerahkan Allah kepada mereka’ (dianggap *X* variabel prediktor).

Mempelajari matematika yang sesuai paradigma taqwa tidak cukup berbekal kemampuan intelektual semata, akan tetapi dibutuhkan secara bersama dengan kemampuan emosional dan spiritual. Pola pikir deduktif dan logis dalam matematika juga bergantung pada kemampuan intuitif imajinatif serta mengembangkan pendekatan rasional empiris dan logis.

Sering kali dijumpai dalam masyarakat umum sebuah pandangan bahwa konsep agama dan matematika tidak memiliki relasi yang setara. Agama yang diekspresikan oleh para pemeluknya di satu sisi cenderung memfokuskan diri pada kegiatan yang bersifat ritual suci dan ukhrawi, sedangkan matematika memiliki corak yang kental. Namun, dalam sejarah dapat dicermati bahwa agama ternyata memiliki peran yang signifikan dalam membangun umatnya untuk mengkaji ilmu matematika lebih mendalam.

2.5.2 Ayat Al-Quran Tentang *Outlier*

وَوُضِعَ الْكِتَابُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيْلَتَنَا مَالِ هَذَا الْكِتَابِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّا أَحْصَاهَا ۗ وَوَجَدُوا مَا عَمِلُوا حَاضِرًا ۗ وَلَا يَظْلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

“Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang yang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya dan mereka berkata : “Aduhai celaka kami, kitab apakah ini yang tidak meninggalkan yang kecil dan tidak (pula) yang besar, melainkan ia mencatat semuanya” dan mereka dapati apa yang telah mereka kerjakan ada (tertulis). Dan Tuhanmu tidak menganiaya seorang juapun” (QS. al-Kahfi/18:49).

Selain menjelaskan tentang pentingnya mengolah data, surat al-Kahfi ayat 49 di atas ini juga menjelaskan tentang data yang menimpang atau data *outlier*. Pada ayat tersebut menjelaskan bahwa kita dalam kehidupan tidak selalu menemui

orang-orang yang benar, tetapi terdapat pula orang-orang yang bersalah. Ada pula diantara mereka yang berpaling dari haluan yang benar. Barang siapa yang beriman kepada Allah dan mentaati-Nya, maka sesungguhnya dia akan menempuh jalan yang akan menyampaikan kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkan dari siksa neraka.

Jika ditelaah ayat di atas menjelaskan suatu penyimpangan, layaknya suatu data yang mengalami penyimpangan dari sekumpulan data. Sehingga dari gambaran di atas dapat diketahui bahwa itulah contoh *outlier* dalam al-Quran.

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

Dalam tafsir *Fi Dzilalil Quran* menjelaskan bahwa sesungguhnya diantara kami (setelah mendengar al-Quran itu) ada golongan menjadi muslim dan ada pula golongan menjadi nyeleweng. Oleh karena itu, barang siapa yang menjadi muslim, maka merekalah orang-orang yang memilih jalan hidayah (Quth, 2008).

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

Setelah diuraikan di atas dapat diambil kesimpulan bahwa yang menjelaskan *outlier* adalah kalimat “*Dan diletakan kitab, lalu kamu akan melihat orang-orang yang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya*” dalam arti *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 pengertiannya yaitu: Yang pertama, *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 sisaannya. Yang kedua, *outlier* adalah suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data yang lainnya (Drape dan Smith, 1998). Dan yang ketiga, *outlier* adalah data yang tidak mengikuti pola umum model (Sembiring, 1995)

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

Dijelaskan dalam sebuah tafsir bahwa diantara hamba-hamba Allah yang hidup dialam semesta ini adalah ada yang muslim dan ada juga yang melakukan penyimpangan. Maksudnya disini adalah mereka melakukan penyimpangan terhadap kebenaran Allah, berarti mereka jauh dari kebenaran-kebenaran Allah (Katsir, 2007).

Dapat diketahui bahwa Allah adalah Dzat yang ahli dari segalanya melebihi ahli-ahli dan pakar-pakar ilmu lainnya. Jadi, jika dibumi Allah ini

terdapat ilmu matematika, maka Allah adalah ahlinya yang paling mengetahui. Dialah Allah Dzat ahli matematika. Dan meskipun di bumi Allah ini terdapat ilmu fisika, biologi, arsitek dan lain-lain, maka Allah paling mengetahui tentang semua itu. Tidak ada yang tidak diketahui oleh Allah. Dan tidak ada yang tidak diketahui Allah. Tidak ada yang tersembunyi bagi Allah sesuatupun yang terjadi di bumi dan bahkan di langit, wallahua'lam (Abdussakir, 2007).



BAB IV

METODE PENELITIAN

3.1 Kerangka Penelitian

Untuk melakukan suatu penelitian perlu adanya kerangka pemikiran sebagai penuntun untuk menjelaskan konsep dari penelitian itu sendiri. Kerangka pemikiran akan memudahkan para pembaca secara jelas dan ringkas mengenai apa yang dilakukan peneliti.

Hal pertama yang dilakukan peneliti adalah mengumpulkan informasi serta data yang bersangkutan dengan *outlier* dan cara penyelesaiannya dengan menggunakan metode MM-Estimasi. Data tersebut kemudian diproses sesuai dengan teori metode MM-Estimasi pada literatur yang tersedia.

Berdasarkan analisis data yang telah dilakukan, maka diharapkan dapat mengidentifikasi permasalahan pada data tersebut. Apabila permasalahannya adalah terdapat *outlier* dalam kumpulan data tersebut, maka peneliti akan melakukan pendeteksian dimana letak keberadaan *outlier* tersebut.

3.2 Metode Pengumpulan Data

Metode pengumpulan data digunakan untuk mengumpulkan data, baik data primer atau sekunder yang diperlukan dalam penelitian. Langkah metode pengumpulan data tersebut harus dilakukan dengan teliti dan bila perlu dilakukan dengan berulang-ulang supaya tidak terjadi kesalahan yang mengakibatkan penelitian tersebut akan gagal dalam melakukan penarikan hasil yang efektif.

Data primer adalah data yang diperoleh atau dikumpulkan oleh peneliti secara langsung dari sumber datangnya. Untuk mendapatkan data primer, peneliti

harus mengumpulkannya secara langsung. Teknik yang dapat digunakan peneliti untuk mengumpulkan data primer adalah dengan cara antara lain observasi, wawancara, diskusi dan penyebaran kuisioner.

Sedangkan yang dimaksud data skunder adalah data yang diperoleh atau dikumpulkan peneliti dari berbagai sumber yang sudah ada seperti halnya jurnal, buku, biro pusat statistik, perpustakaan, media dan lain-lain.

Kali ini peneliti melakukan penelitian dengan menggunakan data skunder yang diperoleh dari hasil riset H. N. Cahya (2010) tentang *survival time* yang mendeskripsikan tentang empat variabel prediktor (x) dan satu variabel respon (y).

Berikut adalah table 3.1 penyajian data yang digunakan dalam penelitian kali ini (H. N. Cahya 2010).

Tabel 3.1 Survival Time

No.	x1	x2	x3	x4	y
1	7.8	65	115	4.3	509
2	5.8	38	72	1.42	80
3	5.1	59	66	1.7	101
4	6.5	73	41	2.01	101
5	7.4	57	83	2.16	204
6	6.7	62	81	2.59	200
7	5.7	46	63	1.91	80
8	3.7	68	81	2.57	127
9	6	67	92	2.5	202
10	3.7	76	94	2.4	203
11	6.3	84	83	4.13	329
12	6.7	51	43	1.86	65
13	7.4	74	68	2.4	217
14	7.7	62	67	3.4	168
15	5.8	83	88	3.95	330
16	7.3	68	74	3.56	215
17	5.6	57	87	3.02	172
18	6	85	28	2.98	87
19	3.7	51	41	1.55	34
20	5.2	52	76	2.85	109

21	6.7	26	68	2.1	70
22	3.4	83	53	1.12	136
23	5.8	96	114	3.95	830
24	5.8	67	86	3.4	220
25	6.3	59	100	2.95	276
26	5.8	61	73	3.5	144
27	5.2	52	86	2.56	181
28	5.8	76	59	2.58	178
29	5.2	54	56	2.71	72
30	11.2	76	90	5.59	574
31	3.2	64	65	0.74	71
32	5.4	58	70	2.64	115
33	5.8	72	93	3.3	295
34	5	59	73	3.5	116
35	8.7	45	23	2.52	58
36	5.3	57	99	2.6	184
37	2.6	74	86	2.05	118
38	5.4	52	88	1.18	148
39	4.8	61	76	2.45	151
40	4.3	8	120	2.85	120
41	5.2	49	72	1.84	95
42	3.4	77	93	1.48	191
43	6.5	40	84	3	123
44	4.5	73	106	3.05	311
45	3.6	28	99	1.3	75
46	8.8	90	88	6.4	483
47	6.7	56	77	2.85	153
48	5.1	67	77	2.86	158
49	8.8	78	72	3.2	313
50	4.8	86	101	4.1	398
51	6.4	85	40	1.21	128
52	6.6	77	46	1.95	124
53	6.4	59	85	2.33	198
54	3.8	82	108	4.55	310

3.3 Analisis Data

Metode penelitian selanjutnya adalah setelah semua data sudah terkumpulkan dan akan diproses, maka hal pertama yang harus dilakukan terlebih

dulu adalah peneliti menganalisa data tersebut. Setelah menganalisa datanya peneliti akan memproses sesuai dengan apa yang akan diteliti. Dalam penelitian kali ini peneliti akan membahas tentang data yang mengandung *outlier*. Setelah data tersebut dinyatakan terdapat *outlier*, maka akan dianalisa dimana letak terdapatnya *outlier*. Kalau *outliernya* terdapat pada arah sumbu x, maka disebut dengan nilai *leverage*. apabila terletak pada arah sumbu y, maka disebut nilai *discrepancy*. Dan apabila terletak pada arah keduanya (x dan y), maka disebut sebagai nilai *influence*.

Setelah semua langkah-langkah diatas selesai, maka data tersebut akan diproses sesuai apa yang akan diteliti. Dan pada penelitian kali ini, peneliti ingin menganalisis tentang ketahanan dan keakuratan metode MM-Estimasi terhadap adanya sebuah *outlier*.

Berikut langkah-langkah detilnya metode penganalisisan penelitian:

6. untuk pengidentifikasian *outlier* dengan nilai *leverage* menggunakan metode *h-matrik*,
7. untuk pengidentifikasian *outlier* dengan nilai *discrepancy* menggunakan metode *externally studentized residual*,
8. untuk nilai *influence* diidentifikasi dengan metode *DFFIT'S* dan *Cook's*,
9. metode *robust* yang digunakan dalam penelitian ini adalah MM-Estimasi untuk menganalisis sekaligus mengatasi *outlier* pada model regresi.

Setelah langkah-langkah diatas selesai dilakukan maka peneliti dapat menarik kesimpulan tentang cocok atau tidak cocok metode MM-Estimasi tersebut untuk mengatasi data yang mengandung *outlier*.

BAB VBAB IV

PEMBAHASAN

Pada bab ini akan dijelaskan tentang bagaimana cara dan tahap-tahap mendeteksi *outlier* pada arah scatterplot x atau scatterplot y atau pada arah keduanya. Kemudian akan dianalisis dengan metode *robust* MM-Estimasi.

Data yang dipakai merupakan data skunder yang diperoleh dari H. N. Cahyadi (2010). Data tersebut merupakan data *survival time* yang berjumlah 54 dengan satu variabel y dan empat variabel x. 10 data pertama akan disajikan dalam Tabel (4.1) seperti di bawah ini dan data selengkapnya akan disajikan dalam Lampiran 1.

Tabel 4.1 *Survival Time*

No.	x1	x2	x3	x4	y
1	7.8	65	115	4.3	509
2	5.8	38	72	1.42	80
3	5.1	59	66	1.7	101
4	6.5	73	41	2.01	101
5	7.4	57	83	2.16	204
6	6.7	62	81	2.59	200
7	5.7	46	63	1.91	80
8	3.7	68	81	2.57	127
9	6	67	92	2.5	202
10	3.7	76	94	2.4	203

4.1 Deteksi *Outlier*

Sebelum menganalisis menggunakan metode *robust*, terlebih dahulu akan diperiksa apakah data tersebut mengandung *outlier* pada arah sumbu x (dengan metode *Leverage*) atau sumbu y (dengan metode *Discrepancy*) atau bahkan pada arah keduanya (dengan metode *Influence*).

4.1.1 Leverage

Seperti yang telah dijelaskan pada bab sebelumnya, bahwa deteksi *outlier* dengan *leverage* adalah memeriksa apakah terdapat *outlier* pada arah x dari setiap pengamatan ke- i dengan melihat nilai h_{ii} dan dengan membandingkan nilai *centroidnya*.

Data yang lebih besar dari nilai *centroid (mean)* variabel independen dan nilai h_{ii} yang melebihi nilai *cutoff* $2(k + 1/n)$ dengan k merupakan banyaknya variabel independen dan n merupakan banyaknya data, maka akan digolongkan sebagai *outlier*. Nilai-nilai *mean (centroid)* dari masing-masing keempat data independen adalah 5.783333 (x_1), 63.42593 (x_2), 77.2037 (x_3), 2.774259 (x_4). Sedangkan nilai *cutoffnya* adalah 0,1852. Untuk data survival time akan ditampilkan 10 data pertama dari nilai *cutoff* dan nilai h -nya pada Tabel (4.2) dan data selengkapnya akan disajikan dalam Lampiran 2.

Table 4.2 deteksi leverage pada data survival time

No	x1	x2	x3	x4	y	cutoff	HI1	Jenis data
1	7.8	65	115	4.3	509	0.1852	0.124131	bukan
2	5.8	38	72	1.42	80	0.1852	0.08127	bukan
3	5.1	59	66	1.7	101	0.1852	0.036983	bukan
4	6.5	73	41	2.01	101	0.1852	0.081727	bukan
5	7.4	57	83	2.16	204	0.1852	0.106676	bukan
6	6.7	62	81	2.59	200	0.1852	0.039845	bukan
7	5.7	46	63	1.91	80	0.1852	0.047382	bukan
8	3.7	68	81	2.57	127	0.1852	0.060138	bukan
9	6	67	92	2.5	202	0.1852	0.055526	bukan
10	3.7	76	94	2.4	203	0.1852	0.07453	bukan

Untuk menentukan nilai yang termasuk *leverage* dapat dihipotesiskan sebagai berikut:

$$H_0: h_{ii} \leq \text{cutoff}, \text{ maka data bukan merupakan leverage}$$

$H_1: h_{ii} \geq cutoff$, maka data merupakan *leverage*

Dari perhitungan didapatkan bahwa data ke-18, 30, 35, 40, 46 mempunyai nilai h_{ii} masing-masing adalah 0.189087, 0.264561, 0.219006, 0.31775, 0.260665 yang mana nilainya lebih besar dari nilai *cutoff*nya yaitu 0.1852. Oleh karena itu, seperti hipotes di atas H_0 ditolak sehingga data-data tersebut terbukti merupakan *outlier* pada arah x (*leverage*).

4.1.2 *Discrepancy*

Nilai *discrepancy* adalah pendeteksian *outlier* pada arah y yang penghitungannya dapat dilakukan dengan dua metode yaitu *internally studentized residuals* dan *externally studentized residuals*. Pemeriksaan *outlier* pada arah y pada penelitian kali ini dilakukan dengan menggunakan metode *externally studentized residuals* (t_i).

Penentuan nilai *outlier* berdasarkan metode *externally studentized residuals* berdasarkan nilai *cutoff*nya yang mengikuti distribusi t dengan $df=n-k-1$. Untuk menentukan nilai tersebut adalah *outlier* diberikan hipotesis jika nilai t_i lebih besar dari nilai t_{tabel} dengan derajat kepercayaan α , maka data tersebut dikategorikan sebagai *outlier*

Untuk data pada Tabel (4.1) di atas tersebut dengan menggunakan metode *discrepancy* menunjukkan bahwa nilai t_{tabel} dengan derajat kepercayaan $\alpha = 0,05$ adalah ± 2.01 dan 10 nilai yang dihasilkan akan ditunjukkan pada Tabel (4.3) di bawah ini dan untuk data selengkapnya akan dilampirkan pada Lampiran 3.

Table 4.3 deteksi *discrepancy* pada data *survival time*

No.	TRES1	t_{tabel}	Jenis data
1	1.0872	2.01	Bukan
2	0.4743	2.01	Bukan
3	0.0746	2.01	Bukan

4	-0.06	2.01	Bukan
5	-0.6346	2.01	Bukan
6	-0.5893	2.01	Bukan
7	0.4706	2.01	Bukan
8	-0.5425	2.01	Bukan
9	-1.2699	2.01	Bukan
10	-0.708	2.01	Bukan

Dari data Tabel (4.3) di atas dengan menggunakan metode *externally studentized residuals* menghasilkan bahwa nilai pada data ke-19 dan ke-23 masing-masing adalah 2.2679 dan 10.0852 yang menunjukkan bahwa nilai data tersebut lebih besar dari nilai t_{tabel} . Oleh karena itu, bisa dikatakan bahwa data ke-19 dan ke-23 adalah *outlier*.

4.1.3 Influence

Metode yang ketiga untuk mendeteksi keberadaan serta letak *outlier* adalah metode *influence*. Seperti yang dijelaskan pada bab dua di atas bahwa metode nilai *influence* adalah kombinasi dari nilai *leverage* dan nilai *discrepancy*. Lebih spesifiknya metode *influence* adalah metode yang mendeteksi keberadaan dan letak *outlier* pada data x dan data y . Metode *influence* tersebut mendeteksi *outlier* dengan cara melihat nilai dari jarak *Cook's* dan *DFFITs*.

Dihipotesiskan bahwa penentuan *outlier* dengan menggunakan *Cook's* adalah jika nilai yang melebihi F_{tabel} dengan $df = (k + 1, n - k - 1)$ dengan derajat kepercayaan $\alpha = 0.05$, maka data tersebut digolongkan sebagai *outlier*. Sedangkan hipotesis untuk menentukan *outlier* dengan melihat nilai *DFFITs* adalah jika nilai yang melebihi $2\sqrt{((K + 1)/n)}$ maka disebut sebagai *outlier*. Dari data Table (3.1) nilai F_{tabel} untuk menentukan berdasarkan *Cook's* adalah 2.31, sedangkan nilai *cutoff* untuk menentukan berdasarkan *DFFITs* adalah ± 0.61 .

Untuk nilai 10 data pertama dengan metode *influence* menggunakan *Cook's* dan *DFITs* akan disajikan dalam Tabel (4.4) di bawah ini dan untuk data selengkapnya akan disajikan dalam Lampiran 4.

Table 4.4 deteksi nilai *influence* pada data *survival time*

No.	<i>COOK1</i>	<i>cutoff</i>	<i>DFIT1</i>	<i>cutoff</i>	Jenis data
1	0.03338	2.31	0.4093	0.61	bukan
2	0.00404	2.31	0.14107	0.61	bukan
3	0.00004	2.31	0.01462	0.61	bukan
4	0.00007	2.31	-0.01791	0.61	bukan
5	0.00974	2.31	-0.21928	0.61	bukan
6	0.00292	2.31	-0.12005	0.61	bukan
7	0.00224	2.31	0.10496	0.61	bukan
8	0.00382	2.31	-0.13724	0.61	bukan
9	0.01873	2.31	-0.30792	0.61	bukan
10	0.00816	2.31	-0.30792	0.61	bukan

Dari data pada Tabel (4.4) di atas menjelaskan bahwa nilai data ke-19, ke-23 dan ke-35 masing-masing adalah 0.95775, 4.29843 dan 0.6478 sehingga data tersebut dikatakan *outlier* karena nilai yang dihasilkan melebihi nilai *cutoff*nya yaitu 0.61.

4.2 MM-Estimasi

MM-Estimasi dimulai dengan mencari S Estimasi yang sangat *robust* dan resisten yang meminimumkan suatu skala residual. Selanjutnya skala residual tetap konstan dan diakhiri dengan menetapkan parameter-parameter regresi menggunakan M Estimasi. MM-Estimasi mempunyai *breakdown point* yang tinggi sama dengan S Estimasi yaitu sebesar 0,5 atau 50%, sehingga MM-Estimasi dapat menjelaskan bahwa banyaknya *outlier* hingga separuh data pengamatan tidak berpengaruh terhadap MM-estimasi. Berikut formula MM-Estimasi dapat dijelaskan seperti di bawah ini:

$$\hat{\beta}_{mm} = \min \Sigma \rho \left(\frac{e_i}{\hat{\sigma}} \right) = \min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=1}^n x_{ij} \beta_j}{\hat{\sigma}} \right)$$

S Estimasi sebagai permulaan dengan nilai high breakdown yang tinggi (50%) dan diakhiri dengan M Estimasi yang membuatnya mempunyai efisiensi yang tinggi. Pada umumnya digunakan fungsi *Tukey Bisquare* baik pada S Estimasi maupun M Estimasi.

Prosedur MM-Estimasi dapat diuraikan sebagai berikut:

1. Mengestimasi koefisien $\hat{\beta}_j^{(1)}$, sehingga diperoleh *residual* $e_i^{(1)}$ yang diambil dari regresi *robust* dengan *high breakdown point*.
2. *Residual* $e_i^{(1)}$ pada langkah pertama digunakan untuk menghitung skala residual M Estimasi, $\hat{\sigma}$ dan dihitung pula bobot awal $w_i^{(1)}$.
3. *Residual* $e_i^{(1)}$ dan skala residual $\hat{\sigma}$ dari langkah kedua digunakan dalam iterasi awal dengan metode WLS untuk menghitung koefisien regresi $\sum_{i=1}^n w_i^{(1)} \left(\frac{e_i^{(1)}}{\hat{\sigma}} \right) x_i = 0$ dimana $w_i^{(1)}$ menggunakan pembobot *Huber* atau *Tukey Bisquare*.
4. Menghitung pembobot baru $w_i^{(1)}$ menggunakan *residual* dari iterasi awal WLS (langkah ketiga)
5. Langkah kedua, ketiga dan keempat diulang (reiterasi dengan skala *residual* tetap konstan) sampai $\sum_{i=1}^n |e_i^{(m)}|$ konvergen, yaitu selisih β_j^{m+1} dengan

β_j^m kurang dari 10^{-4} atau sampai mendekati 0 (m = banyaknya iterasi), maka iterasi dihentikan.

Penerapan metode MM-Estimasi pada data table (4.1) di atas menghasilkan model:

$$\hat{y} = -353.6101 + 25.8923x_1 + 2.4334x_2 + 2.7282x_3 + 0.4754x_4$$

Dengan metode MM-Estimasi, *R-square* yang dihasilkan adalah 94,1% dan menghasilkan estimasi residual sebesar 20,96. Hal tersebut menyatakan bahwa keakuratan metode MM-Estimasi tidak terpengaruh dengan adanya *outlier*, sehingga dapat diputuskan metode MM-Estimasi sangat cocok digunakan apabila bila dalam suatu data penelitian terdapat *outlier*.

4.3 Kajian *Outlier* dalam Al-Quran

Outlier dapat diartikan sebuah data pencilan dari kebanyakan data atau bisa diartikan juga bahwa *outlier* adalah sebuah data yang tidak mengikuti ritme sebuah alur dari kebanyakan data yang ada. Seperti yang telah didefinisikan oleh Barnett dan Lewis (1994), *outlier* adalah data yang muncul tidak konsisten dengan sisa data. Secara umum *outlier* dapat diartikan data yang tidak mengikuti pola umum model, atau dapat dikatakan sebagai data yang menyimpang. Dalam kehidupan, *outlier* dapat dikatakan sebagai sesuatu yang menyimpang dalam kebenaran.

Menyimpang dari kebenaran berhubungan dengan amal perbuatan manusia. Amal perbuatan manusia terbagi menjadi dua yaitu amal perbuatan baik

dan amal perbuatan yang buruk. Pada al-Quran dijelaskan pada surat Fussilat/41:46, yaitu:

مَنْ عَمِلَ صَالِحًا فَلِنَفْسِهِ ۖ وَمَنْ أَسَاءَ فَعَلَيْهَا ۚ وَمَا رَبُّكَ بِظَلَّامٍ لِلْعَبِيدِ ﴿٤٦﴾

“Barangsiapa yang mengerjakan amal shaleh maka (pahalanya) untuk dirinya sendiri dan barang siapa mengerjakan perbuatan jahat, maka (dosanya) untuk dirinya sendiri; dan sekali-kali tidaklah Rabb-mu menganiaya hamba-hambaNya. (QS. Fussilat/41:46).

Dijelaskan dalam al-Quran surat Fussilat ayat ke-46 ini *barang siapa mengerjakan amal saleh* yaitu amal yang diperintahkan Allah dan Rasul-Nya maka *pahala* (manfaatnya) *untuk dirinya sendiri* dan *barang siapa berbuat jahat* maka (dosa dan hukumannya) *menjadi tanggungan dirinya sendiri* dalam ayat ini terdapat dorongan untuk mengerjakan kebaikan dan meninggalkan keburukan, adanya akibat dari amal yang dilakukan, bahwa seseorang tidak dapat memikul dosa orang lain, *dan tuhanmu sama sekali tidak menzalimi hamba-hamba-Nya* seperti memikul kepada hamba dosa-dosa diluar dosa mereka (Marwan bin Musa, Tafsir Hidayatul Insan).

Outlier merupakan salah satu faktor yang dapat mempengaruhi pendugaan parameter. Yang dapat mengakibatkan data tidak konsisten. Ada sebab, maka ada akibat. Seperti halnya perbuatan manusia, semua amalan baik dan buruk akan mendapatkan balasannya. Hal ini di bahas dalam surat al-Mukmin/40:40, yaitu:

مَنْ عَمِلَ سَيِّئَةً فَلَا يُجْزَىٰ إِلَّا مِثْلَهَا ۖ وَمَنْ عَمِلَ صَالِحًا مِّن ذَكَرٍ أَوْ أُنْثَىٰ وَهُوَ مُؤْمِنٌ ۚ فَأُولَٰئِكَ يَدْخُلُونَ الْجَنَّةَ يُرْزَقُونَ فِيهَا بِغَيْرِ حِسَابٍ ﴿٤٠﴾

“Barangsiapa mengerjakan perbuatan jahat, maka dia tidak akan dibalasi melainkan sebanding dengan kejahatan itu. Dan barangsiapa mengerjakan amal shaleh baik laki-laki maupun perempuan sedang ia dalam keadaan beriman. Maka mereka akan masuk surga, mereka diberi rezki di dalamnya tanpa hisab.” (QS. al-Mukmin/40:40).

Dari surat al-Mukmin/40:40 ini dijelaskan bahwa *barangsiapa mengerjakan perbuatan jahat, maka dia akan dibalas sebanding dengan kejahatan itu. Dan barangsiapa mengerjakan amal shaleh baik yang berkaitan dengan hati, lisan maupun anggota badan, baik laki-laki maupun perempuan sedangkan dia dalam keadaan beriman, maka mereka akan masuk surga, mereka diberi rezeki dalamnya tak terhingga, Allah akan memberikan rezeki kepada mereka yang tidak dicapai oleh amal mereka* (Marwan bin Musa, Tafsir Hidayatul Insan.



BAB V

PENUTUP

5.1 Kesimpulan

Metode pendeteksian *outlier* yang dibahas adalah metode *leverage*, *discrepancy*, *influence*. Sedangkan metode *robust* yang dipakai dalam penelitian ini adalah metode MM-Estimasi yang mana metode tersebut menggabungkan antara *high breakdown* dan *efficiency* yang tinggi pula.

Untuk *leverage* pendeteksian *outlier* pada arah x , nilai *cutoff* yang telah ditentukan adalah $2(k + 1)/n$ untuk data yang banyak $n > 15$. Nilai *discrepancy* pendeteksian *outlier* pada arah y yang penghitungannya dapat dilakukan dengan dua metode yaitu *internally studentized residuals* dan *externally studentized residuals*. Nilai *influence* merupakan kombinasi dari nilai *leverage* dan nilai *discrepancy*. Metode *influence* tersebut mendeteksi *outlier* dengan cara melihat nilai dari *Cook's* dan *DFFITs*.

Metode MM-Estimasi menggabungkan antara *high breakdown* dan *efficiency* yang tinggi, untuk *high breakdown*nya mencapai 50% dan sifat keefficiencynya mencapai 95%. Dari hasil *output* data Table (4.1) didapatkan bahwa persamaan metode MM-Estimasi adalah:

$$\hat{y} = -353.6101 + 25.8923x_1 + 2.4334x_2 + 2.7282x_3 + 0.4754x_4$$

Dengan metode MM-Estimasi, *R-square* yang dihasilkan adalah 94,1% dan menghasilkan estimasi *residual* sebesar 20,96. Hal tersebut menyatakan bahwa keakuratan metode MM-Estimasi tidak terpengaruh dengan adanya *outlier*,

sehingga dapat diputuskan metode MM-Estimasi sangat cocok digunakan apabila dalam suatu data penelitian terdapat *outlier*.

5.2 Saran

Pada tugas akhir ini hanya digunakan metode MM-Estimasi untuk mengatasi pengaruh *outlier*. Oleh karena itu, untuk penelitian yang lain supaya digunakan metode yang lebih banyak seperti halnya *Least Trimmed Squares (LTS) Estimate*, *M Estimate*, *S Estimate*, dll.



DAFTAR PUSTAKA

- Abdussakir. 2007. *Ketika Kyai Mengajar Matematika*. Malang: UIN Press.
- Algifari, 1997. *Analisis Regresi Teori Kasus dan Solusi*. Yogyakarta: BPFE.
- Cohen, J.. 2003. *Applied Multiple Regression/Correlation Analysis For The Behavioral Sciences*. New Jercey: Lawrence Erlbaum Associate.
- Draper, N., dan Smith, H.. 1998. *Analisis Regresi Terapan*. Jakarta: PT. Gramedia Pustaka Utama.
- Sembiring, R.K.. 1995. *Analisis Regresi*. Bandung: ITB.
- Soemarti. 2007. *Pencilan (Outlier)*. Makalah Statitika FMIPA Universitas Padjadjaran. Bandung. Tersedia: [http://resources.unpad.ac.id/unpad-content/aplounds/publikasi_dosen/Outlier\(Pencilan\).pdf](http://resources.unpad.ac.id/unpad-content/aplounds/publikasi_dosen/Outlier(Pencilan).pdf).
- Hasan, I.. 2002. *Pokok-pokok Materi Metodologi Penelitian dan Aplikasinya*. Jakarta: Ghalia Indonesia.
- Hasan, I.. 2002. *Pokok-pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta: Bumi Askara.
- Quth, S.. 2008. *Tafsir Fidzitalil Qur'an*. Jakarta: Gema Press.
- Seber, G.. 2007. *Linier Regression Analipsis*. New Zewland: Intersince.
- Wibisono, Y.. 2005. *Metode Statistik*. Yogyakarta: Gajah Mada University Press.
- Yitnosumarto, S.. 1990. *Dasar-dasar Statistik*. Jakarta: CV. Rajawali.

Lampiran I
Tabel 4.1 *Survival Time*

No.	x1	x2	x3	x4	y
1	7.8	65	115	4.3	509
2	5.8	38	72	1.42	80
3	5.1	59	66	1.7	101
4	6.5	73	41	2.01	101
5	7.4	57	83	2.16	204
6	6.7	62	81	2.59	200
7	5.7	46	63	1.91	80
8	3.7	68	81	2.57	127
9	6	67	92	2.5	202
10	3.7	76	94	2.4	203
11	6.3	84	83	4.13	329
12	6.7	51	43	1.86	65
13	7.4	74	68	2.4	217
14	7.7	62	67	3.4	168
15	5.8	83	88	3.95	330
16	7.3	68	74	3.56	215
17	5.6	57	87	3.02	172
18	6	85	28	2.98	87
19	3.7	51	41	1.55	34
20	5.2	52	76	2.85	109
21	6.7	26	68	2.1	70
22	3.4	83	53	1.12	136
23	5.8	96	114	3.95	830
24	5.8	67	86	3.4	220
25	6.3	59	100	2.95	276
26	5.8	61	73	3.5	144
27	5.2	52	86	2.56	181
28	5.8	76	59	2.58	178
29	5.2	54	56	2.71	72
30	11.2	76	90	5.59	574
31	3.2	64	65	0.74	71
32	5.4	58	70	2.64	115
33	5.8	72	93	3.3	295
34	5	59	73	3.5	116
35	8.7	45	23	2.52	58
36	5.3	57	99	2.6	184
37	2.6	74	86	2.05	118

38	5.4	52	88	1.18	148
39	4.8	61	76	2.45	151
40	4.3	8	120	2.85	120
41	5.2	49	72	1.84	95
42	3.4	77	93	1.48	191
43	6.5	40	84	3	123
44	4.5	73	106	3.05	311
45	3.6	28	99	1.3	75
46	8.8	90	88	6.4	483
47	6.7	56	77	2.85	153
48	5.1	67	77	2.86	158
49	8.8	78	72	3.2	313
50	4.8	86	101	4.1	398
51	6.4	85	40	1.21	128
52	6.6	77	46	1.95	124
53	6.4	59	85	2.33	198
54	3.8	82	108	4.55	310

Lampiran II

Table 4.2 deteksi *leverage* pada data *survival time*

No.	x1	x2	x3	x4	y	<i>cutoff</i>	h_{ij}	Jenis data
1	7.8	65	115	4.3	509	0.1852	0.124131	bukan
2	5.8	38	72	1.42	80	0.1852	0.08127	bukan
3	5.1	59	66	1.7	101	0.1852	0.036983	bukan
4	6.5	73	41	2.01	101	0.1852	0.081727	bukan
5	7.4	57	83	2.16	204	0.1852	0.106676	bukan
6	6.7	62	81	2.59	200	0.1852	0.039845	bukan
7	5.7	46	63	1.91	80	0.1852	0.047382	bukan
8	3.7	68	81	2.57	127	0.1852	0.060138	bukan
9	6	67	92	2.5	202	0.1852	0.055526	bukan
10	3.7	76	94	2.4	203	0.1852	0.07453	bukan
11	6.3	84	83	4.13	329	0.1852	0.065277	bukan
12	6.7	51	43	1.86	65	0.1852	0.080764	bukan
13	7.4	74	68	2.4	217	0.1852	0.08157	bukan
14	7.7	62	67	3.4	168	0.1852	0.050458	bukan
15	5.8	83	88	3.95	330	0.1852	0.059973	bukan
16	7.3	68	74	3.56	215	0.1852	0.038603	bukan
17	5.6	57	87	3.02	172	0.1852	0.027375	bukan
18	6	85	28	2.98	87	0.1852	0.189087	<i>outlier</i>
19	3.7	51	41	1.55	34	0.1852	0.151355	bukan
20	5.2	52	76	2.85	109	0.1852	0.027375	bukan
21	6.7	26	68	2.1	70	0.1852	0.124505	bukan
22	3.4	83	53	1.12	136	0.1852	0.141251	bukan
23	5.8	96	114	3.95	830	0.1852	0.153731	bukan
24	5.8	67	86	3.4	220	0.1852	0.02792	bukan
25	6.3	59	100	2.95	276	0.1852	0.055787	bukan
26	5.8	61	73	3.5	144	0.1852	0.055338	bukan
27	5.2	52	86	2.56	181	0.1852	0.031417	bukan
28	5.8	76	59	2.58	178	0.1852	0.042994	bukan
29	5.2	54	56	2.71	72	0.1852	0.083994	bukan
30	11.2	76	90	5.59	574	0.1852	0.264561	<i>outlier</i>
31	3.2	64	65	0.74	71	0.1852	0.106039	bukan
32	5.4	58	70	2.64	115	0.1852	0.028826	bukan
33	5.8	72	93	3.3	295	0.1852	0.033969	bukan
34	5	59	73	3.5	116	0.1852	0.92267	bukan
35	8.7	45	23	2.52	58	0.1852	0.219006	<i>outlier</i>
36	5.3	57	99	2.6	184	0.1852	0.047335	bukan
37	2.6	74	86	2.05	118	0.1852	0.105797	bukan
38	5.4	52	88	1.18	148	0.1852	0.061503	bukan
39	4.8	61	76	2.45	151	0.1852	0.027146	bukan
40	4.3	8	120	2.85	120	0.1852	0.31775	<i>outlier</i>
41	5.2	49	72	1.84	95	0.1852	0.038149	bukan

42	3.4	77	93	1.48	191	0.1852	0.132737	bukan
43	6.5	40	84	3	123	0.1852	0.068236	bukan
44	4.5	73	106	3.05	311	0.1852	0.068236	bukan
45	3.6	28	99	1.3	75	0.1852	0.148072	bukan
46	8.8	90	88	6.4	483	0.1852	0.260665	<i>outlier</i>
47	6.7	56	77	2.85	153	0.1852	0.026771	bukan
48	5.1	67	77	2.86	158	0.1852	0.027048	bukan
49	8.8	78	72	3.2	313	0.1852	0.123695	bukan
50	4.8	86	101	4.1	398	0.1852	0.09949	bukan
51	6.4	85	40	1.21	128	0.1852	0.174638	bukan
52	6.6	77	46	1.95	124	0.1852	0.085847	bukan
53	6.4	59	85	2.33	198	0.1852	0.050524	bukan
54	3.8	82	108	4.55	310	0.1852	0.177734	bukan



Lampiran III

Table 4.3 deteksi *discrepancy* pada data *survival time*

No.	TRES1	t_{tabel}	Jenis data
1	1.0872	2.01	Bukan
2	0.4743	2.01	Bukan
3	0.0746	2.01	Bukan
4	-0.06	2.01	Bukan
5	-0.6346	2.01	Bukan
6	-0.5893	2.01	Bukan
7	0.4706	2.01	Bukan
8	-0.5425	2.01	Bukan
9	-1.2699	2.01	Bukan
10	-0.708	2.01	Bukan
11	-0.2344	2.01	Bukan
12	0.6928	2.01	Bukan
13	-0.6398	2.01	Bukan
14	-0.896	2.01	Bukan
15	-0.1704	2.01	Bukan
16	-0.8199	2.01	Bukan
17	-0.5737	2.01	Bukan
18	-0.1818	2.01	Bukan
19	2.2679	2.01	<i>Outlier</i>
20	-0.2622	2.01	Bukan
21	0.8161	2.01	Bukan
22	0.9634	2.01	Bukan
23	10.0852	2.01	<i>Outlier</i>
24	-0.6061	2.01	Bukan
25	-0.2639	2.01	Bukan
26	-0.5839	2.01	Bukan
27	0.3227	2.01	Bukan

Lampiran IV

Table 4.4 deteksi nilai *influence* pada data *survival time*

No.	<i>COOK1</i>	<i>cutoff</i>	<i>DFIT1</i>	<i>cutoff</i>	Jenis data
1	0.03338	2.31	0.4093	0.61	bukan
2	0.00404	2.31	0.14107	0.61	bukan
3	0.00004	2.31	0.01462	0.61	bukan
4	0.00007	2.31	-0.01791	0.61	bukan
5	0.00974	2.31	-0.21928	0.61	bukan
6	0.00292	2.31	-0.12005	0.61	bukan
7	0.00224	2.31	0.10496	0.61	bukan
8	0.00382	2.31	-0.13724	0.61	bukan
9	0.01873	2.31	-0.30792	0.61	bukan
10	0.00816	2.31	-0.30792	0.61	bukan
11	0.00078	2.31	-0.20093	0.61	bukan
12	0.00852	2.31	-0.06194	0.61	bukan
13	0.00736	2.31	-0.19068	0.61	bukan
14	0.00857	2.31	-0.20654	0.61	bukan
15	0.00038	2.31	-0.04304	0.61	bukan
16	0.00543	2.31	-0.1643	0.61	bukan
17	0.00188	2.31	-0.09625	0.61	bukan
18	0.00157	2.31	-0.08777	0.61	bukan
19	0.16915	2.31	0.95775	0.61	<i>Outlier</i>
20	0.00065	2.31	-0.05641	0.61	Bukan
21	0.01907	2.31	0.30775	0.61	Bukan
22	0.03058	2.31	0.39071	0.61	Bukan
23	1.20946	2.31	4.29843	0.61	<i>Outlier</i>
24	0.00214	2.31	-0.10272	0.61	Bukan
25	0.00084	2.31	-0.06415	0.61	Bukan
26	0.00405	2.31	-0.14133	0.61	Bukan
27	0.00069	2.31	0.05813	0.61	Bukan
28	0.00003	2.31	0.01183	0.61	Bukan
29	0.00265	2.31	0.114	0.61	Bukan
30	0.07152	2.31	0.59798	0.61	Bukan
31	0.00772	2.31	0.1951	0.61	Bukan
32	0.00037	2.31	-0.04252	0.61	Bukan
33	0.00023	2.31	-0.03372	0.61	Bukan
34	0.00449	2.31	-0.14865	0.61	Bukan
35	0.08309	2.31	0.6478	0.61	<i>Outlier</i>
36	0.00898	2.31	-0.21172	0.61	Bukan
37	0.01367	2.31	-0.26027	0.61	Bukan
38	0.00154	2.31	-0.08696	0.61	Bukan
39	0.00005	2.31	0.01628	0.61	Bukan
40	0.03792	2.31	0.43278	0.61	Bukan
41	0.00029	2.31	0.03755	0.61	Bukan
42	0.01016	2.31	-0.22386	0.61	Bukan
43	0.00358	2.31	-0.13276	0.61	Bukan
44	0.00013	2.31	-0.02564	0.61	Bukan

45	0.01024	2.31	0.22459	0.61	Bukan
46	0.00848	2.31	-0.20406	0.61	Bukan
47	0.00205	2.31	-0.10065	0.61	Bukan
48	0.00153	2.31	-0.08668	0.61	Bukan
49	0.00903	2.31	-0.21096	0.61	Bukan
50	0.00356	2.31	0.13228	0.61	Bukan
51	0.00251	2.31	-0.11097	0.61	Bukan
52	0.00236	2.31	-0.10755	0.61	Bukan
53	0.00234	2.31	-0.10732	0.61	Bukan
54	0.03622	2.31	-0.42486	0.61	Bukan



RIWAYAT HIDUP

Lukmanul Hakim dilahirkan di Gresik pada tanggal 30 April 1991, anak pertama dari pasangan bapak Moh. Sholeh dan ibu Munawaroh. Pendidikan dasar ditempuh di kampung halamannya di Madrasah Ibtida'iyah Mojopetung Dukun Gresik yang ditamatkan pada tahun 2003. Pada tahun yang sama melanjutkan pendidikan menengah pertama di Madrasah Tsanawiyah Al-Falahiyah Mojopetung Dukun Gresik sampai pada tahun 2006. Kemudian melanjutkan pendidikan menengah atas di Madrasah Aliyah 07 Sunan Drajat Lamongan. dan menamatkan pendidikan tersebut pada tahun 2009. Pendidikan berikutnya ditempuh di Universitas Islam Negeri Maulana Malik Ibrahim Malang dengan mengambil Jurusan Matematika Fakultas Sains dan Teknologi.

