DETEKSI *OUTLIER* PADA MODEL REGRESI *ROBUST* DENGAN METODE *LEAST TRIMMED SQUARE* (LTS)

SKRIPSI

Oleh:
Dewi Ratnasari
NIM. 08610078

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2014

DETEKSI *OUTLIER* PADA MODEL REGRESI *ROBUST* DENGAN METODE *LEAST TRIMMED SQUARE* (LTS)

SKRIPSI

Diajukankepada:
FakultasSainsdanTeknologi
Univeritas Islam NegeriMaulana Malik Ibrahim Malang
untukMemenuhiSalah SatuPersyaratan
dalamMemperolehGelarSarjanaSains (S.Si)

Oleh:
Dewi Ratnasari
NIM. 08610078

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2014

DETEKSI *OUTLIER* PADA MODEL REGRESI *ROBUST* DENGAN METODE *LEAST TRIMMED SQUARE* (LTS)

SKRIPSI

Oleh:
Dewi Ratnasari
NIM. 08610078

Telah Diperiksa dan Disetujui untuk Diuji

Tanggal: 07 April 2014

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

Mengetahui,

Ketua Jurusan Matematika

Dr. Abdussakir, M.Pd
NIP. 19751006200312 1 001

DETEKSI *OUTLIER* PADA MODEL REGRESI *ROBUST* DENGAN METODE *LEAST TRIMMED SQUARE* (LTS)

SKRIPSI

Oleh:
Dewi Ratnasari
NIM. 08610078

TelahDipertahankan di DepanDewanPenguji
danDinyatakanDiterimasebagaiSalah SatuPersyaratan
untukMemperolehGelarSarjanaSains (S.Si)
Tanggal: 10 April 2014

Susunan Dewan Penguji			Tanda Tangan
1.	Penguji Utama	: Abdul Aziz, M.Si NIP. 19760318 100604 1 002	( )
2.	Ketua Penguji	: EvawatiAlisah, M.Pd NIP. 19720604 199903 2 001	( )
3.	Sekretaris Penguji	:Dr. Sri Harini, M.Si NIP. 19731014 200112 2 002	( )
4.	Anggota Penguji	: Dr. H. Ahmad Barizi, M.A NIP. 19731212 199803 1 001	( )

Mengesahkan,
Ketua Jurusan Matematika

Dr. Abdussakir, M.Pd
NIP.19751006200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: DEWI RATNASARI
NIM	: 08610078
Jurusan	: Matematika
Fakultas	: Sains dan Teknologi
Judul	: Deteksi *Outlier* Pada Model Regresi *Robust* dengan Metode *Least Trimmed Square* (LTS)

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 07 April 2014

Yang membuat pernyataan,

Dewi Ratnasari

NIM. 08610078

MOTTO

‘Karena sesungguhnya sesudah kesulitan itu ada kemudahan, dan sesungguhnya sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain, dan hanya kepada Tuhanmulah kamu berharap’

(QS. Alam Nasyrah ayat 5-8)

PERSEMBAHAN

Karya ini penulis persembahkan untuk

orang-orang yang telah memberikan arti bagi hidup penulis

Dengan pengorbanan, kasih sayang dan ketulusannya.

Kepada kedua orang tua penulis ayahanda dan ibunda yang paling berjasa dalam hidup penulis dan selalu menjadi motivator dan penyemangat dalam setiap langkah penulis untuk terus berproses menjadi insan kamil. Terima kasih atas ketulusan dan keihlasannya dalam memberikan kasih sayang selama ini sehingga menjadikan hidup penulis begitu indah dan lebih berarti, penulis persembahkan karya ini kepada ayahanda dan ibunda.

Hanya do'a dan harapan yang terucap:

Semoga Allah SWT memberikan kekuatan dan kemampuan kepada penulis

untuk bisa mewujudkan apa yang ayahanda dan ibunda titipkan selama ini.

"Amin Ya Robbal Alamin"

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahirobbil 'alamin, segala puji syukur ke hadirat Allah SWT atas limpahan rahmat, taufiq, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Sholawat serta salam semoga senantiasa tercurahkan kepada nabi besar Muhammad SAW sebagai *Uswatun Hasanah* dalam meraih kesuksesan di dunia dan akhirat.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan *jazakumullahu ahsanal jaza'* kepada semua pihak yang telah membantu selesainya skripsi ini. Ucapan terima kasih ini penulis sampaikan 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 yang telah memberikan arahan dan pengalaman yang berharga.
4. Dr. Sri Harini, M.Si dan Dr. Ahmad Barizi, M.A, selaku dosen pembimbing skripsi, yang telah memberikan banyak arahan dan pengalaman yang berharga.
5. Abdul Aziz, M.Si dan Evawati Alisah, M.Pd, selaku dosen penguji skripsi, terima kasih telah memberikan masukan-masukan yang berharga dan bermanfaat untuk penulisan skripsi ini.

6. Mohammad Jamhuri, M.Si selaku dosen wali yang memberikan arahan dan nasihat yang berharga.
7. Seluruh dosen Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah membantu dalam menyelesaikan skripsi ini.
8. Ayahanda dan Ibunda yang telah mencurahkan cinta dan kasih sayang teriring do'a, restu, motivasi, dan materi, sehingga penulis selalu optimis dalam menggapai salah satu kesuksesan hidup.
9. Kakak penulis, Sulasmani, S.Pd, Giono, dan keponakan penulis Muhammad August dan Muhammad Augan yang telah memberikan dukungan, do'a, motivasi serta keceriaan selama ini.
10. Teman-teman terbaik penulis, Azizizah Noor Aini, S.Si, Irhasyah Fitrotul Afifi, S.Si, Ummu Aiman Chabasiyah, S.Si, dan Aulia Dewi Farisky, S.Si yang telah memberikan keceriaan tersendiri dalam hidup penulis. Terima kasih atas segala pengalaman berharga dan kenangan terindah yang telah terukir.
11. Sahabat-sahabat penulis seperjuangan di Jurusan Matematika tercinta Ahmad Munawir, Nur Miftah, dan Lailatul Maghfiroh serta seluruh teman-teman Jurusan Matematika khususnya angkatan 2008 dan 2009 yang berjuang bersama-sama untuk mencapai kesuksesan yang diimpikan. Terima kasih atas segala pengalaman berharga dan kenangan terindah yang telah terukir.

12. Sahabat, kakak, dan teman segalanya penulis Rusiaji Pramudita, S.M yang senantiasa ada di saat suka dan duka. Terima kasih banyak atas doa, nasihat, dan motivasi yang selalu diberikan.
13. Kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, yang tidak bisa disebutkan satu per satu.

Terakhir, penulis sadar bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, kritik dan saran konstruktif dari para pembaca yang budiman sangat penulis harapkan demi perbaikan dan kebaikan skripsi ini.

Semoga skripsi ini dapat bermanfaat dan berguna bagi kita semua, terutama bagi diri penulis sendiri. *Amin ya Robbal 'Alamin…*

Malang, April 2014

Penulis,

DAFTAR ISI

BAB III : PEMBAHASAN

BAB IV : PENUTUP

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Ratnasari, Dewi. 2014. **Deteksi *Outlier* pada Model Regresi *Robust* dengan Metode *Least Trimmed Square* (LTS).** Skripsi. Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (I) Dr. Sri Harini, M.Si
(II) Dr.H.Ahmad Barizi, M.A

Kata Kunci: *Outlier*, Regresi *Robust*, *Least Trimmed Square*

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*.

Regresi *robust* merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Pada penelitian ini model yang digunakan dalam mendeteksi *outlier* adalah model regresi *robust* .Dimana dari model tersebut akan diestimasi dengan metode *least trimmed square.*

Metode *least trimmed square* merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat *h* residual (fungsi objektif). Metode *least trimmed square* lebih *resistance* terhadap data yang mengandung *outlier* dibanding dengan metode *ordinary least square*.

Dari hasil penelitian didapatkan estimasi parameter model regresi *robust* dengan metode *least trimmed square* yang memenuhi syarat *unbias*. Sehingga $\hat{\beta}_{LTS}$ yang didapat sudah dapat digunakan untuk mengatasi *outlier* pada model regresi *robust*.

ABSTRACT

Ratnasari, Dewi. 2014. **Outlier Detection on Robust Regression Role with Least Trimmed Square (LTS).** Thesis. Department of Mathematics. Faculty of Science and Technology. State Islamic University Maulana Malik Ibrahim Malang.
Advisor: (I) Dr. Sri Harini, M.Si
(II) Dr.H.Ahmad Barizi, M.A

Kata Kunci: Outlier, Robust Regression, Least Trimmed Square

In general, outlier can be interpreted as a data that did not follow a common pattern in the model or data out of the model and not within the area of the confidence interval. Outlier is one of the factors that can affect the estimation of parameters in linear regression models. To determine whether outliers affect the estimation of parameters in a linear regression model we estimate parameters of a linear regression model containing outlier and the estimation of these parameters and apply to the data containing outlier.

Robust regression is an important way to analize data that influenced by outlier. Therefore a robust or outlier resistant model is obtained. In this study the model that is used in detecting outlier is robust regression model. In which from that model the regression model will be estimated with least trimmed square method.

Least trimmed square method is an estimation method of parameter robust regression to minimalize the number of residual up to h (objective function). Least trimmed square method is more resistant for data outlier than ordinary least square method.

From the result of the study we obtain estimation parameter model of robust regression with least trimmed square method that meets unbias property. Therefore the resulting $\hat{\beta}_{LTS}$ can be used to overcome outlier on robust regression model.

الملخّص

رتنسرئ ,ديري.2014. **كشف اوتلير عند نموج الافحدار القوي باستخدام طريقة *Least Trimmed Square***.
البحث الجامعي. الشعبة الرياضيات لكلية العلوم والتكنولوجيا بجامعة الإسلامية الحكوميةمولان مالك ابراهيم مالانج.
المشرف : د.سري هاريني الماجستير
د. أ. بارزي الماجستي

الكلمة الرئيسية : أوتلير، الانحدار القوي، *Least Trimmed Square*

عامة، أنّ أوتلير تعني البيانات التي لاتتبع أشكال عامة أو من أشكال البيانات سوى أشكالها ولاتوجدفي فاصل الثقة. أوتلير من العوامل التي يتأثرتقدير الممثلةفي نموذج الإنحدار القطي. لمعرفة أم كانت أوتلير أن تتأثر تقدير الممثلة ل الإنحدار القطي تُستخدَم تقديرممثلة في التي لها أوتلير و أن تطبّق حاصلها على البيانات التي لها الأوتلير

الانحدار القوي هو أداة مهمة لتحليل البيانات التي تتأثر القيم المتطرفة حتى أن النماذج المحصولة قوية أو مقاومة عن آوتلير .في هذه البحث لستخدمنا الإنحدار القوي لكشف أوتلير. سيقدر هذا النموذج بـ *Least Trimmed Square*.

طريقة *Least Trimmed Square* هي طريقة المعلمات للإنحدار القوي لتقليل h (دالة الهدف). طريقة LTS اكثر مقاومة للبيانات الأوتلير من الطريقة OLS.

من نتائج هذا البحث تقدير المعلمة لنموذج الإنحدار القوي باستخدام *Least Trimmed Square* التي مؤهلة عن شرط *unbias*. حتى نحصل على $\hat{\beta}_{LTS}$ التي يمكننا آن نستخدمها لتجا وز آوتلير في نموذج الإنحدا القوي.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika adalah cabang matematika yang berkaitan dengan pengumpulan data, pengolahan data, penyajian data, analisis data, dan penarikan kesimpulan. Suatu kegiatan utama dalam statistika adalah pengumpulan data. Dalam masalah mengumpulkan data yaitu mencatat atau membukukan data, Al-Qur'an juga membicarakannya. Perhatikan Al-Qur'an surat Al-Kahfi ayat 49.

وَوُضِعَ ٱلْكِتَٰبُ فَتَرَى ٱلْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَٰوَيْلَتَنَا مَالِ هَٰذَا
ٱلْكِتَٰبِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحْصَىٰهَا ۚ وَوَجَدُوا۟ مَا عَمِلُوا۟ حَاضِرًا ۗ
وَلَا يَظْلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

Artinya: "Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang 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."

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 الكتب yang berarti mencatat, pada statistik langkah awal adalah mencatat terlebih dahulu data apa yang dibutuhkan. Setelah melakukan pencatatan barulah dapat mengolah hasil dari pencatatan data tersebut. Pada ayat di atas juga terdapat kata عملوآ yang berarti data. Jadi dalam statistik terdapat kegiatan mencatat dan mengumpulkan data baru setelah itu dapat mengolah data tersebut dan memberikan kesimpulan.

Selain kegiatan mengumpulkan data, statistika juga melakukan analisis untuk mencari hubungan antara data atau variabel. Analisis regresi bertujuan untuk mengetahui pola dan hubungan antara variabel respon serta mengestimasi parameter regresi dalam model. Salah satu metode estimasi parameter pada model regresi adalah *Ordinary Least Square* (OLS). Metode estimasi ini dapat digunakan ketika asumsi model regresi linier terpenuhi. Akan tetapi dalam kenyataan di lapangan terkadang data yang dihasilkan tidak memenuhi model regresi linier. Salah satunya adalah munculnya *outlier* pada model tersebut. Jika pada model regresi linier yang mengandung *outlier* diselesaikan dengan metode *Ordinary Least Square* (OLS), maka hasil estimasi yang diperoleh adalah bias dan tidak efisien. Salah satu alternatif untuk mengatasi kelemahan metode OLS adalah dengan menggunakan metode regresi *robust*. Regresi *robust* digunakan untuk mengestimasi parameter model regresi linier pada data yang mengandung *outlier*. Sehingga dihasilkan model yang *robust* atau mampu mendeteksi dan menyelesaikan *outlier*.

Penelitian yang mengkaji tentang data yang mengandung *outlier* telah dilakukan oleh Isnaini (2000). Penelitian Isnaini menggunakan Metode Kuadrat Terkecil (*Least Squares*) untuk menduga parameter model, padahal pada data yang mengandung pencilan (*Outlier*). Metode Kuadrat Terkecil mempunyai kelemahan yaitu sangat rentan terhadap keberadaan data pencilan (*Outlier*). Karena alasan tersebut penulis menerapkan regresi *robust* dengan menggunakan metode *Least Trimmed Squares* (LTS) dalam menduga parameter model. Keunggulan metode *robust* adalah model yang diperoleh akan bersifat tegar atau tidak terpengaruh oleh adanya titik-titik *outlier*.

Berdasarkan latar belakang di atas, maka penelitian ini akan membahas tentang "Deteksi *Outlier* pada Model Regresi *Robust* dengan Metode *Least Trimmed Square* (LTS)".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah bagaimanakah prosedur regresi *robust* pada data Pendapatan Pedagang Kaki Lima dengan menggunakan metode *Least Trimmed Square* (LTS) dengan meminimalkan *outlier* ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui prosedur dan model regresi robust serta hasil mendeteksi yang meminimalkan *outlier* pada data Pendapatan Pedagang Kaki Lima dengan menggunakan metode *Least Trimmed Square* (LTS).

1.4 Batasan Masalah

Berdasarkan rumusan masalah dan tujuan penelitian yang telah disebutkan di atas, maka batasan masalah yang diberikan adalah model regresi yang digunakan pada penelitian ini adalah model regresi linier. Metode *Ordinary*

Least Square (OLS) dan *Least Trimmed Square* (LTS) hanya untuk data Pendapatan Pedagang Kaki Lima.

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 pendeteksian *outlier* pada model regresi *robust* dengan metode *Least Trimmed Square* (LTS).

2. Bagi Pembaca

   Dapat dijadikan sebagai tambahan referensi bagi mahasiswa matematika dalam memahami khususnya ilmu statistik 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 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 dari orang lain) sebagai literatur untuk mengumpulkan data-data dan informasi (Hasan, 2002).

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. Dimana data lapang hanyalah untuk melengkapi penggunaan prosedur regresi *robust* dengan metode *Least Trimmed Square* (LTS).

Adapun langkah-langkah dalam penelitian ini adalah:

1. Estimasi parameter model regresi *robust* dengan metode *Least Trimmed Square* (LTS).
2. Menentukan sifat estimasi parameter model regresi *robust*.
   a. Parameter Tak Bias (Parameter *Unbias*)
   b. Efisien
   c. Konsisten
3. Aplikasi data
   a. Mendeskripsi data dengan mencari grafik plot menggunakan program MINITAB
   b. Deteksi *outlier*
4. Persamaan regresi *robust*.
5. Interpretasi data

1.7 Sistematika Penulisan

Untuk mempermudah dalam 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, metode penelitian, dan sistematika penulisan.

BAB II : Tinjauan Pustaka

Pada bab ini akan dibahas tentang konsep-konsep (teori-teori) yang mendukung bagian pembahasan.

BAB III : Pembahasan

Pada bab ini dijelaskan prosedur regresi *robust* pada data Pendapatan Pedagang Kaki Lima dengan menggunakan metode *Least Trimmed Square* (LTS) dengan meminimalkan *outlier*.

BAB IV : Penutup

Pada bab ini dibahas tentang kesimpulan dari hasil pembahasan dan saran.

BAB II

TINJAUAN 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 sekali antar generasi. Hal ini dijelaskan Galton pada fakta yang memperlihatkan adanya kecenderungan mundurnya (*regresi*) 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 estimasi atau prediksi dan rata-rata nilai variabel tergantung dengan diketahuinya nilai variabel bebas (Algifari, 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. 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 $X_1, X_2, \ldots, X_i$ adalah variabel-variabel bebas dan Y adalah variabel terikat, maka terdapat hubungan fungsional antara X dan Y, di mana variasi dari X akan diiringi pula oleh variasi dari Y. Analisis regresi adalah teknik analisis yang mencoba menjelaskan bentuk hubungan antara peubah-peubah yang mendukung sebab akibat. Prosedur analisisnya didasarkan atas distribusi probabilitas bersama peubah-peubahnya. Bila hubungan ini dapat dinyatakan dalam persamaan matematika maka dapat memanfaatkan untuk keperluan-keperluan lain misalnya peramalan. Tujuan utama dari analisis regresi adalah mendapatkan dugaan (ramalan) dari suatu variabel dengan menggunakan variabel lain yang diketahui. Analisis regresi mempunyai dua jenis pilihan yaitu regresi linier dan regresi non linier.

Wiibisono (2005) menyatakan bahwa untuk menguji model analisis regresi, terdapat empat langkah yang menguji model analisis regresi, antara lain:

1. Menentukan estimasi parameter dari model regresi linier,
2. Uji normalitas data,
3. Menguji asumsi homoskedatisitas, dan
4. Uji asumsi multikolinieritas.

2.2 Regresi Linier

Regresi merupakan suatu alat ukur untuk mengukur ada atau tidaknya 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 antar variabel tersebut dapat dilakukan pendugaan suatu variabel berdasarkan variabel lain melalui persamaan yang dihubungkan tersebut (Algifari, 1997).

Model regresi linear secara umum dapat dinyatakan dengan

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e \qquad (2.1)$$

dimana :

y = variabel terikat

x = variabel bebas

$\beta_0, \beta_1, \ldots, \beta_k$ = parameter model

e = *error*

2.3 Model Regresi Linier dalam Pendekatan Matriks

Sembiring (1995) menyatakan bahwa model regresi linier sederhana dapat digeneralisasikan menjadi lebih dari satu atau k variabel. Persamaan bagi model regresi linier dengan k variabel adalah sebagai berikut

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \ldots + \beta_k x_k + e \qquad (2.2)$$

Bila pengamatan $y, x_1, x_2, \ldots, x_k$ dinyatakan masing-masing dengan $y_i, x_{i1}, x_{i2}, \ldots, x_{ik}$ dan *error* e_i, maka persamaanya adalah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} + e_i \qquad (2.3)$$

dimana $i = 1,2,3, \ldots, n$. Dan dinotasikan dalam bentuk matriks menjadi

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \qquad (2.4)$$

misalkan:

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

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Maka secara ringkas persamaan 2.4 dapat ditulis sebagai berikut

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{2.5}$$

dimana:

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

$\boldsymbol{X}$: vektor peubah bebas ukuran $n \times (k+1)$

$\boldsymbol{\beta}$: vektor parameter ukuran $(k+1) \times 1$

$\boldsymbol{e}$: vektor galat ukuran $n \times 1$

dengan asumsi $e \sim N(\mu, \sigma^2)$

2.4 *Outlier*

Metode diagnosis *error* dalam regresi digunakan untuk memeriksa kelayakan berbagai asumsi yang mendasari proses pemodelan serta untuk menemukan keanehan yang terkandung dalam data. *Error* dikatakan sebagai *outlier* jika mempunyai nilai mutlak yang jauh lebih besar dari pada *error* lainnya dan bisa jadi terletak 3 atau 4 simpangan baku dari rata-ratanya.

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).

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 dikemukakan oleh Soemarti (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. *Outlier* merupakan nilai ekstrim dari suatu pengamatan. Volume *error* dari suatu *outlier* sama dengan selisih antara kuantitas (ukuran) observasi dan kuantitas prediksi untuk observasi yang ke – *i*. Dari persamaan 2.3 diperoleh persamaan *outlier* sebagai berikut

$$e_i = y_i - \hat{y}_i = y_i - x_i\beta \qquad (2.6)$$

dimana

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_k x_{ik} = x_i\beta \qquad (2.7)$$

2.5 Mendeteksi *Outlier*

Terdapat beberapa cara yang dapat digunakan untuk mengidentifikasi adanya *outlier*, antara lain :

2.5.1 Nilai Pengaruh (*Leverage Value*)

Nilai pengaruh digunakan untk mengidentifikasi adanya *outlier* pada peubah prediktor (X) 2.1 mempunyai k+1 parameter, dan matriks *hat* (H) didefinisikan sebagai :

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}} \qquad (2.8)$$

Untuk $i = 1,2,\dots,n$, maka nilai pengaruh h_{ii} dari nilai $X(x_{i1}, x_{i2}, \dots, x_{ik})$ didefinisikan sebagai elemen ke-*i* dari diagonal dari matriks *hat* (H) adalah

$$h_{ii} = x_i(X^T X)^{-1} x_i^T \qquad (2.9)$$

dimana:

$$x_i^T = [1 \quad x_{11} \quad x_{12} \quad x_{13} \quad \cdots \quad x_{ik}] \qquad (2.10)$$

Pengamatan ke-*i* dianggap sebagai *outlier* jika

$$h_{ii} > 2\bar{h} = 2\frac{\sum_{i=1}^{n} h_{ii}}{n} \qquad (2.11)$$

Karena jumlah dari h_{ii} untuk *i* sama dengan 1 sampai *n* sama dengan jumlah koefisien model regresi maka persamaan 2.11 dapat dituliskan menjadi :

$$h_{ii} > 2\,\frac{(k+1)}{n} \qquad (2.12)$$

2.5.2 *Studentizied Deleted Residual* ($\boldsymbol{TRES}$)

Uji statistik *Studentizied Deleted Residual* ($TRES$) digunakan untuk memeriksa adanya pencilan pada peubah respon Y. Hipotesis yang digunakan dalam pemeriksaan adanya *outlier* adalah :

H_0 : pengamatan ke-*i* bukan merupakan *outlier*

H_1 : pengamatan ke-*i* merupakan *outlier*

Persamaan yang digunakan untuk menghitung TRES dari setiap pengamatan ke-*i* adalah

$$|TRES|_i = \frac{d_i}{s(d_i)} \qquad (2.13)$$

dimana:

d_i = $Y_i - \hat{Y}_{i(i)}$

$s(d_i)$ = simpangan baku dari $d_i = \sqrt{\frac{MSE_{(i)}}{1-h_{ii}}}$

$MSE_{(i)}$ = kuadrat tengah galat dari hasil analisis regresi baru setelah membuang pengamatan ke-*i*

i = 1, 2, 3, ….,n

$\hat{Y}_{i(i)}$ = $b_0^{(i)} + b_1^{(i)}X_{i1} + b_2^{(i)}X_{i2} + \cdots + b_k^{(i)}X_{ik}$ adalah penduga dari Y_i, yang dihitung dengan menggunakan penduga $b_0^{(i)}, b_1^{(i)}, b_2^{(i)}, \ldots, b_k^{(i)}$ untuk data tanpa pengamatan ke-*i*.

Kriteria pengujian yang dilandasi keputusan adalah

1. Jika nilai $|TRES|_i \leq t^{\frac{a}{2}}_{n-k-2}$, maka H_0 diterima

2. Jika nilai $|TRES|_i > t^{\frac{a}{2}}_{n-k-2}$, maka H_0 ditolak

2.5.3 Mendeteksi *Outlier* Berpengaruh

Setelah dilakukan indetifikasi terhadap *outlier* maka langkah selanjutnya adalah mengidentifikasi apakah *outlier* tergolong pengamatan berpengaruh atau pengamatan tidak berpengaruh. Pengamatan berpengaruh adalah pengamatan yang berpengaruh besar terhadap persamaan regresi. Ada cara untuk mendeteksi pengamatan berpengaruh, yaitu DFFITS (*The Difference in Fit Statistics*). DFFITS dari suatu pengamatan ke-i merupakan ukuran pengaruh dari pengamatan ke-i pada nilai duga Y $(\hat{Y}_i)$. Sehingga DFFITS digunakan untuk mendeteksi pengamatan yang berpengaruh terhadap $\hat{Y}_i$. Hipotesis yang digunakan untuk memeriksa apakah pengamatan ke-i termasuk pengamatan berpengaruh adalah :

H_0 : pengamatan ke-i bukan merupakan pengamatan berpengaruh

H_i : pengamatan ke-i merupakan pengamatan berpengaruh

Untuk menghitung nilai DFFITS dari setiap pengamatan ke-i, maka persamaan yang digunakan adalah

$$(DFFITS)_i = \left[\frac{d_i}{s(d_i)}\right]\left[\frac{h_{ii}}{1-h_{ii}}\right]^{\frac{1}{2}} \quad (2.14)$$

Kriteria pengujian yang melandasi keputusan adalah

1. Jika nilai $|DFFITS|_i \leq 2\sqrt{\frac{k+1}{n}}$, maka H_0 diterima

2. Jika nilai $|DFFITS|_i > 2\sqrt{\frac{k+1}{n}}$, maka H_0 ditolak

2.6 Regresi Robust

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi dari *error* tidak normal atau adanya beberapa *outlier* yang berpengaruh pada model. Metode ini merupakan alat penting untuk menganalisa data yang dipengaruhi oleh *outlier* sehingga dihasilkan model yang *robust* atau *resistance* terhadap *outlier*. Suatu estimasi yang resistan adalah relatif tidak terpengaruh oleh perubahan besar pada bagian kecil data atau perubahan kecil pada bagian besar data.

Prosedur *robust* ditujukan untuk mengakomodasi adanya keanehan data, sekaligus meniadakan identifikasi adanya data *outlier* dan juga bersifat otomatis dalam menanggulangi data *outlier*. Beberapa metode estimasi dalam regresi robust diantaranya M-*Estimation*, *Least Trimmed Square* (LTS), MM *estimation*, S *estimation*, dan *Least Mean Square* (Seber, 2007).

Model regresi pada regresi *robust* adalah sebagai berikut

$$Y_i = X_i^T\beta + e_i \tag{2.15}$$

dimana

Y_i = mempunyai ordo (n×1)

X_i^T = mempunyai ordo (n×(k+1))

β = mempunyai ordo ((k+1)×n)

e_i = mempunyai ordo (n×1)

2.7 *Least Trimmed Square*

Least Trimmed Square (LTS) merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat *h* residual (fungsi objektif):

$$\sum_{i=1}^{h} e_{(i:n)}^2 \tag{2.16}$$

untuk,

$$h = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right] \tag{2.17}$$

dimana :

$e_{(i)}^2$: Kuadrat *error* yang diurutkan dari terkecil ke terbesar.

$$e_{(1)}^2 < e_{(2)}^2 < e_{(3)}^2 < \cdots < e_{(i)}^2 < \cdots < e_{(h)}^2 < \cdots < e_{(n)}^2$$

n: Banyaknya pengamatan

k: banyaknya parameter regresi

Jumlah *h* pada persamaan 2.17 menunjukkan sejumlah subset data dengan kuadrat fungsi objektif terkecil. Nilai tersebut akan mempengaruhi besarnya nilai *breakdown* (ε_n). Nilai *breakdown* adalah suatu nilai yang menunjukkan proporsi terkecil dari data yang terpengaruh *outlier* yang dapat menyebabkan penduga bernilai jauh berbeda dengan penduga dari data yang tidak terpengaruh *outlier*. Persamaan yang digunakan untuk menghitung nilai *breakdown* adalah

$$\varepsilon_n = \frac{n-h}{n} \tag{2.18}$$

Dari persamaan 2.18, maka dapat dihitung nilai duga *preliminary scale error* yang merupakan akar dari MSE dari metode LTS dan dihitung menggunakan persamaan

$$S_{(LTS)} = d_{(h,n)}\sqrt{\frac{1}{h}\sum_{i=1}^{h} e_{[i]}^2(\hat{\beta}_{(LTS)})} \qquad (2.19)$$

dimana :

$e_{[i]}^2(\hat{\beta}_{(LTS)})$ = jumlah kuadrat *error* menggunakan penduga LTS

h = konstanta pemotongan

$d_{(h,n)} = \frac{1}{\sqrt{1\frac{2n}{hc_{h,n}\phi(1/c_{h,n})}}}$, dengan $c_{h,n} = \frac{1}{\Phi^{-1}(\frac{h+n}{2n})}$

n = banyaknya pengamatan

ϕ = fungsi sebaran normal

Φ^{-1} = invers fungsi sebaran komulatif normal baku

Untuk menduga parameter model dengan metode LTS digunakan algoritma *basic resampling*. Secara rinci algoritma LTS adalah sebagai berikut:

1. Mengambil m subset contoh yang berisi $k+1$ pengamatan berdasarkan kombinasi banyaknya parameter $k+1$ dari n banyaknya pengamatan C_{k+1}^{n}.

$$m = C_{k+1}^{n} = \frac{n!}{(k+1)!(n-(k+1))!} \qquad (2.20)$$

2. Membentuk model untuk semua subset J_v yang terbentuk dimana $v = 1,2,\dots,m$

$$y = x\beta \qquad (2.21)$$

dimana:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, x = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}$$

3. Mendefinisikan nilai h untuk menghasilkan tingkat ke-*robust*-an yang tinggi

dimana nilai h sebesar $\left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right]$ atau $\left[\frac{n}{2}\right] + \left[\frac{(k+2)}{2}\right]$ kemudian menghitung nilai *breakdown* ε_n^*.

$$\varepsilon_n^* = \frac{n-h}{n}$$

4. Mengevaluasi model regresi yang menggunakan intersep melalui proses *adjustment* untuk setiap subset contoh yang terbentuk dengan langkah-langkah sebagai berikut:

   a. Menduga koefisien regresi β untuk setiap subset contoh menggunakan persamaaan:

   $$\hat{\beta} = x^{-1}y \tag{2.22}$$

   b. Menghitung nilai penduga *error* tanpa menggunakan

   $\beta_0, e_i = y_i - \hat{y}_i, i = 1,2, \dots, n.$

   c. Mengurutkan nilai e_i, sehingga $e_{[1]} < e_{[2]} < \dots < e_{[n]}$.

   d. Membentuk kelas interval masing-masing berisi h pengamatan.

   e. Menghitung nilai intersep yang baru $\hat{\beta}_0 = \bar{e}_{[v]}$ yang diperoleh dari nilai rata-rata kelas interval yang mempunyai jumlah kuadrat simpangan $(sd)_v$ yang terkecil.

   $$Min\ (sd)_v = Min\ \sum_{i=1}^{h}(e_{[i]} - \bar{e})^2 \tag{2.23}$$

   f. Menyatakan penduga koefisien regresi dengan nilai intersep yang baru, sehingga didapat model regresi yang baru untuk setiap subset contoh.

5. Menentukan penduga LTS $\hat{\beta}_{(LTS)}$ berdasarkan nilai Q yang minimum untuk setiap $J_v; v; 1,2, \dots, m$.

6. Menghitung *preliminary scale error* seperti pada persamaan 2.19.

2.8 Keakuratan Model

Untuk mengetahui ukuran keakuratan model regresi dapat dilakukan dengan beberapa prosedur, yaitu:

2.8.1 Koefisien Determinasi Terkoreksi

Koefisien determinasi yang memperhitungkan banyaknya peubah yang memperjelas dalam model disebut koefisien determinasi terkoreksi (R_{adj}^2). Koefisien determinasi ini telah terkoreksi terhadap derajat bebas (*db*) masing-masing jumlah kuadrat atau dapat dikatakan juga bahwa koefisien ini telah terkoreksi oleh keragaman totalnya.

Koefisien determinasi terkoreksi (R_{adj}^2) berkisar antara $0 \leq R_{adj}^2 \leq 1$. Semakin besar nilai R_{adj}^2 maka semakin baik penduga model regresi. Koefisien determinasi terkoreksi (R_{adj}^2) dirumuskan dalam persamaan 2.24 (Drapper dan Smith, 1992).

$$\left(R_{adj}^2\right) = 1\frac{\frac{JK_{galat}}{(n-k-1)}}{\frac{JK_{total}}{(n-1)}} = 1 - \frac{\frac{\sum_{i=1}^{n}\hat{e}_i^2}{(n-k-1)}}{\frac{\sum_{i=1}^{n}(Y_i-\hat{Y})^2}{(n-1)}}$$

$$= 1 - \frac{\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{(n-k-1)}}{\frac{\sum_{i=1}^{n}(Y_i-Y)^2}{(n-1)}} = 1 - \frac{(n-1)}{(n-k-1)}\left(\frac{\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i-Y)^2}\right)$$

$$= 1 - \frac{(n-1)}{(n-k-1)}\left(1 - \frac{JK_{regresi}}{JK_{total}}\right) = 1 - \frac{(n-1)}{(n-k-1)}(1-R^2) \quad (2.24)$$

dimana:

R_{adj}^2 = koefisien determinasi terkoreksi

R^2 = koefisien determinasi

n = banyaknya pengamatan

k = banyaknaya peubah prediktor

Secara umum interpretasi dari koefisien determinasi terkoreksi (R^2_{adj}) adalah proporsi dari keragaman total Y yang dijelaskan oleh keragaman peubah prediktor X_j.

2.8.2 Nilai Tengah Kuadrat *Error*

Nilai tengah kuadrat *error* merupakan suatu ukuran ketepatan perhitungan dengan mengkuadratkan setiap *error* untuk setiap penduga dalam sebuah kumpulan data dan kemudian memperoleh rata-rata atau nilai tengah jumlah kuadrat tersebut. Persamaan MSE adalah

$$MSE = \hat{\sigma}^2_{OLS}$$

$$= \frac{JK_{error}}{db} = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{(n-k-1)}$$

$$= \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{(n-k-1)} \qquad (2.25)$$

Semakin kecil nilainya, maka semakin baik kecocokan suatu model dengan data karena nilai penduga dari Y semakin mendeteksi nilai sebenarnya. Sedangkan MSE dari metode LTS merupakan kuadrat dari *premilinary scale error estimate* seperti pada persamaan 2.19 (Sembiring, 1995).

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

Statistika adalah cabang matematika yang berkaitan dengan pengumpulan data, pengolahan data, analisis data, dan penarikan kesimpulan. Kegiatan utama dalam statistika adalah pengumpulan data, hal ini dibicarakan Al-Qur'an dalam Surat Al- Kahfi ayat 49.

وَوُضِعَ ٱلۡكِتَٰبُ فَتَرَى ٱلۡمُجۡرِمِينَ مُشۡفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَٰوَيۡلَتَنَا مَالِ هَٰذَا ٱلۡكِتَٰبِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحۡصَىٰهَاۚ وَوَجَدُواْ مَا عَمِلُواْ حَاضِرًاۗ وَلَا يَظۡلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

Artinya :" Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang 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."

Pada penggalan ayat tersebut terdapat kata الكتب yang berarti mencatat, pada statistik langkah awal adalah mencatat terlebih dahulu data apa yang dibutuhkan. Selain terdapat kandungan ayat yang berarti mencatat, pada ayat tersebut juga terkandung kata yang berarti bentuk-bentuk data. Bentuk data terkadang nilainya ada yang besar dan terkadang bernilai kecil. Setelah melakukan pencatatan barulah dapat mengolah hasil dari pencatatan data tersebut.

2.9.1 Ayat Al-Qur'an tentang Analisis Regresi

Al-Qur'an merupakan kitab Allah SWT yang di dalamnya terkandung ilmu-ilmu Allah SWT, untuk mendapatkan ilmu tersebut perlu mengkaji Al-Qur'an secara mendalam. Al-Qur'an surat Al-Baqarah ayat 2-3 dapat digunakan untuk analisis regresi dengan cara mempartisinya (membagi) dan hasil partisian ayat tersebut dimisalkan dengan sebuah variabel, yaitu :

ذَٰلِكَ ٱلۡكِتَٰبُ لَا رَيۡبَۛ فِيهِۛ هُدًى لِّلۡمُتَّقِينَ ﴿٢﴾ ٱلَّذِينَ يُؤۡمِنُونَ بِٱلۡغَيۡبِ وَيُقِيمُونَ ٱلصَّلَوٰةَ وَمِمَّا رَزَقۡنَٰهُمۡ يُنفِقُونَ ﴿٣﴾

Artinya : *"Kitab (Al Qur'an) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertaqwa. (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebagian rezeki yang kami anugerahkan kepada mereka."*

Apabila kedua ayat tersebut dipartisi, maka diperoleh sebanyak dua bagian, yaitu :

هُدًى لِّلْمُتَّقِينَ .... (*Y*)

وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ............... (*X*)

Dalam ayat tersebut dijelaskan bahwa tidak ada keraguan di dalam Kitab suci Al-Qur'an. Al-Qur'an ini juga merupakan petunjuk bagi mereka yang *bertaqwa*, dianggap *(Y)* variabel *respon*. Sedangkan kriteria *taqwa* itu adalah gabungan dari orang-orang yang mempunyai karakter '*beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebagian rezeki yang di anugerahkan Allah kepada mereka'* dianggap *(X)* variabel *prediktor*.

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

Seringkali 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 membangunkan umatnya dalam tidur panjangnya untuk mengkaji ilmu matematika lebih mendalam.

2.9.2 Ayat Al-Qur'an tentang *Outlier*

وَوُضِعَ ٱلْكِتَٰبُ فَتَرَى ٱلْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَٰوَيْلَتَنَا مَالِ هَٰذَا ٱلْكِتَٰبِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحْصَىٰهَا ۚ وَوَجَدُوا۟ مَا عَمِلُوا۟ حَاضِرًا ۗ وَلَا يَظْلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

Artinya : "Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang 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."

Selain menjelaskan tentang pentingnya mengolah data, Surat Al-Kahfi ayat 49 juga menjelaskan tentang data yang menyimpang 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 beriman kepada Allah dan mentaati-Nya sesungguhnya dia telah menempuh jalan yang akan menyampaikannya kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkan dari siksa neraka. Jika ditelaah ayat di atas mejelaskan suatu penyimpangan, 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*, bisa menyebabkan 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.

Menurut Sayyid Quth (2008) dalam tafsirnya *Fi Dzilalil Qur'an* menjelaskan bahwa sesungguhnya di antara 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 diletakkanlah kitab, lalu kamu akan melihat orang-orang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya*" 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 (Drape dan Smith 1998).
3. *Outlier* adalah data yang tidak mengikuti pola umum model (Sembiring, 1995).

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 terletak sangat jauh diantara data dalam model tersebut.

Sedangkan menurut tafsir Ibnu Katsir (2007) dijelaskan bahwa diantara hamba-hamba Allah yang hidup dialam semesta ini adalah ada yang muslim ada juga yang melakukan penyimpangan. Maksudnya disini adalah mereka melakukan penyimpangan terhadap kebenaran Allah. Berarti mereka jauh dari kebenaran-kebenaran Allah.

Dapat diketahui bahwa Allah SWT adalah Dzat 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 Dzat 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 (Abdussakir, 2007).

BAB III

PEMBAHASAN

Pada bab ini akan dibahas tentang deteksi *outlier* pada model regresi *robust* dengan metode *Least Trimmed Square* (LTS) untuk mendeteksi adanya data yang mengandung *outlier* tersebut.

3.1. Estimasi Parameter Model Regresi *Robust*

Sebelum mengestimasi model regresi *robust* dengan LTS, maka terlebih dahulu mengestimasi model regresi liniernya menggunakan metode OLS. Dari persamaan 2.3 dan 2.5, didapatkan persamaan jumlah kuadrat *error* sebagai berikut.

$$\begin{aligned} s(\beta) &= \sum_{i=1}^{n} e_i^2 \\ &= e^T e \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= (Y^T - \beta^T X^T)(Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \end{aligned} \tag{3.1}$$

Setelah didapatkan persamaan 3.1 tersebut selanjutnya dicari nilai parameter β dengan cara meminimumkan jumlah kuadrat *error*, dari persamaan kuadrat terkecil tersebut dengan turunan terhadap β dan disamadengankan nol dengan cara sebagai berikut:

$$\frac{dS}{d\beta} = \frac{d(Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta)}{d\beta} = 0$$

$$= \frac{d(Y^TY - (Y^TX\beta)^T - \beta^TX^TY + \beta^TX^TX\beta)}{d\beta} = 0$$

$$= \frac{d(Y^TY - \beta^TX^TY - \beta^TX^TY + \beta^TX^TX\beta)}{d\beta} = 0$$

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

$$= -2X^TY + 2X^TX\beta = 0$$

$$2X^TX\beta = 2X^TY$$

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY \quad (3.2)$$

Jadi estimasi dari parameter $\hat{\beta}_{OLS}$ adalah persamaan 3.2 yaitu :

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$$

3.1.1 Estimasi Parameter dengan LTS

Prinsip dari estimasi LTS adalah meminimumkan jumlah kuadrat *error* dari pengamatan sebanyak h yang dipilih dari C_h^n subset data berukuran h, dimana $h \subset n$. Model regresi *robust* untuk mengestimasi β adalah $Y_i = X_i^T\beta + e_i$.

$$\hat{\beta}_{LTS} = \underset{\beta}{\arg\ min} \sum_{i=1}^{h} e_i^2$$

dimana $e_i = Y_i - X_i^T\beta$

Sebelum mengestimasi parameter $\hat{\beta}_{LTS}$ dari model regresi dicari nilai kuadrat *error*nya sebagai berikut:

$$e = Y - X^T\beta$$

$$e^Te = (Y - X^T\beta)^T(Y - X^T\beta)$$

$$= (Y^T - \beta^TX)(Y - \beta^TX)$$

$$= Y^TY - Y^TX^T\beta - \beta^TXY + \beta^TXX^T\beta \quad (3.3)$$

dengan

$e^T e = S$

Sehingga

$S = \quad Y^T Y - Y^T X^T \beta - \beta^T XY + \beta^T XX^T \beta$

Untuk mengestimasi parameter $\hat{\beta}$ yang dinotasikan dengan β, adalah dengan menurunkan persamaan 3.5 terhadap β dan disamadenganken nol.

Diturunkan terhadap β, sebagai berikut

$$\frac{dS}{d\beta} = \frac{d(Y^T Y - Y^T X^T \beta - \beta^T XY + \beta^T XX^T \beta)}{d\beta} = 0$$

$$= \frac{d(Y^T Y - (Y^T X^T \beta)^T - \beta^T XY + \beta^T XX^T \beta)}{d\beta} = 0$$

$$= \frac{d(Y^T Y - \beta^T XY - \beta^T XY + \beta^T XX^T \beta)}{d\beta} = 0$$

$$= \frac{d(Y^T Y - 2\beta^T XY + \beta^T XX^T \beta)}{d\beta} = 0$$

$$= -2XY + 2XX^T \beta = 0$$

$$\hat{\beta}_{LTS} = (XX^T)^{-1} XY \qquad (3.4)$$

Jadi estimasi dari parameter $\hat{\beta}_{(LTS)}$ adalah persamaan 3.4 yaitu :

$$\hat{\beta}_{LTS} = (XX^T)^{-1} XY$$

3.2. Menentukan Sifat-Sifat Estimasi Parameter

3.2.1 Parameter Tak Bias (Parameter *Unbias*)

Tak bias parameter merupakan nilai harapan dari selisih antara nilai estimasi dan nilai yang sebenarnya. Maka estimator β dikatakan estimator tak bias karena $E(\hat{\beta}_{LTS}) = \beta$.

Di sisi lain dari permasalahan yang dirumuskan parameter estimasi adalah:

$$
\begin{aligned}
E(\hat{\beta}_{LTS}) &= E[(XX^T)^{-1}XY] \\
&= E[(XX^T)^{-1}X]E(Y) \\
&= (XX^T)^{-1}X(X^T\beta) \\
&= (XX^T)^{-1}XX^T\beta \\
&= I\beta \\
&= \beta \qquad (3.5)
\end{aligned}
$$

Dari persamaan 3.5 diperoleh $E(\hat{\beta}_{LTS}) = \beta$ maka $\hat{\beta}_{LTS}$ merupakan estimator tak bias.

3.2.2 Efisien

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

Perhatikan bahwa :

$$
\begin{aligned}
Var\,(\hat{\beta}_{LTS}) &= E\left[\left(\hat{\beta}_{LTS} - E(\beta)\right)\left(\hat{\beta}_{LTS} - E(\hat{\beta}_{LTS})\right)^T\right] \\
&= E\left[(\hat{\beta}_{LTS} - \beta)(\hat{\beta}_{LTS} - \beta)^T\right]
\end{aligned}
$$

karena

$$\hat{\beta}_{LTS} = (XX^T)^{-1}XY$$

$$= (XX^T)^{-1}X(X^T\beta + e)$$

$$= (XX^T)^{-1}XX^T\beta + Xe$$

$$= I\beta + Xe$$

$$= \beta + Xe$$

$$\hat{\beta}_{LTS} - \beta = Xe$$

maka

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

$$= E[(Xe)(Xe)^T]$$

$$= E[Xee^TX^T]$$

$$= XE(e)E(e^T)X^T$$

$$= XE(ee^T)X^T$$

$$= XX^T\sigma^2 I$$

$$= XX^T\sigma^2 \qquad (3.6)$$

Sehingga $Var\,(\hat{\beta}_{LTS}) = XX^T\sigma^2$ harus sekecil mungkin agar $\hat{\beta}_{LTS}$ efisien.

3.2.3 Konsisten

Estimator yang konsisten adalah

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

Dari persamaan 3.5 diperoleh $E(\hat{\beta}_{LTS}) = \beta$, maka

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

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

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

$$= (0)\left(\hat{\beta}_{LTS} - \beta\right)^T$$

$$= 0 \tag{3.7}$$

Dari persamaan 3.7 diperoleh

$$E\left[\left(\hat{\beta}_{LTS} - E(\beta)\right)\left(\hat{\beta}_{LTS} - E\left(\hat{\beta}_{LTS}\right)\right)^T\right] = 0,$$

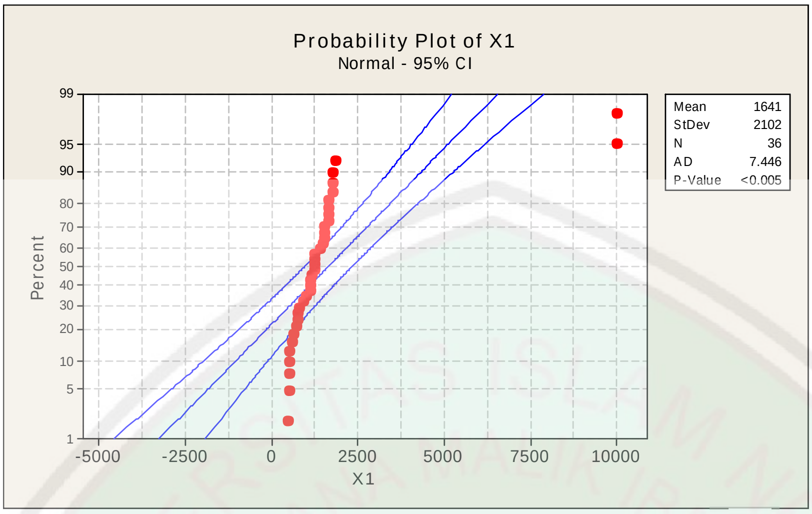
maka $\hat{\beta}_{LTS}$ merupakan estimator yang konsisten.

3.3. Aplikasi pada Estimasi Parameter Model Regresi *Robust*

3.3.1 Deskripsi Data

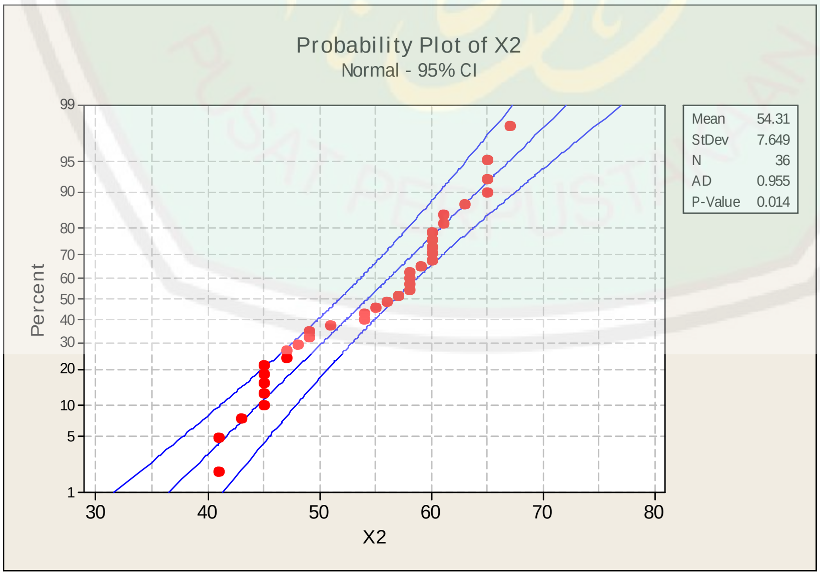
Data yang digunakan dalam skripsi ini dari Fatmawati (2008). Data ini diambil dari 36 pedagang kaki lima seperti pada lampiran 1. Data ini adalah data faktor-faktor yang mempengaruhi pendapatan bersih pedagang kaki lima (PKL) yang dipengaruhi oleh modal PKL (X_1), jam kerja PKL (X_2), dan lama usaha PKL (X_3). Data yang digunakan dalam penelitian ini adalah data sekunder yang mengandung *outlier* dan sudah memenuhi asumsi analisis regresi dengan uji probabilitas.

Dari data (lampiran 1) tersebut dapat dibuat grafik dengan menggunakan MINITAB 16 dan mengasumsikan $\alpha = 0,05$. Hasilnya adalah sebagai berikut :



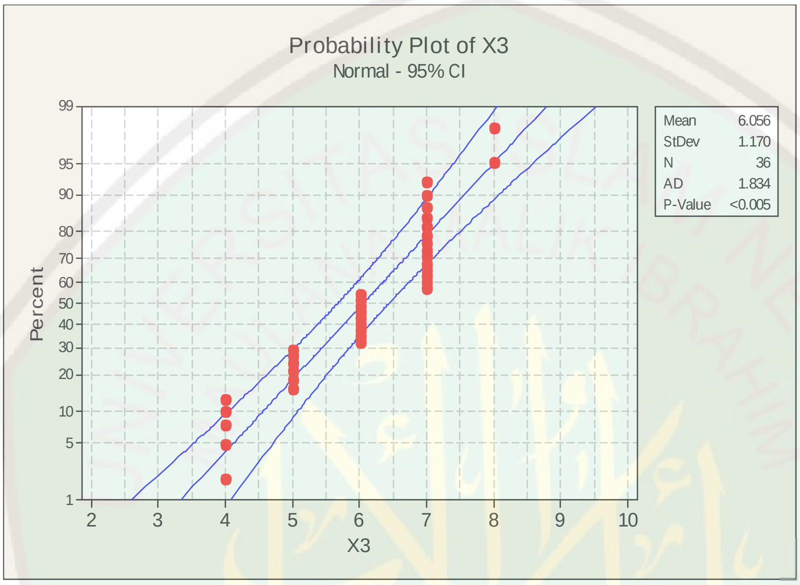
Gambar 3.1 Grafik Modal PKL

Gambar 3.1 di atas merupakan gambar modal (X_1) dari 36 Pedagang Kaki Lima, dengan nilai rata-rata sebesar 1641 dan standart deviasi 2102. Dari gambar 3.1 juga dapat dilihat bahwa sebaran data tidak normal, karena p-value $< \alpha$ (0,05). Jika dilihat sekilas ada beberapa data yang letaknya jauh dari garis normal.



Gambar 3.2 Grafik Jam Kerja PKL

Gambar 3.2 di atas merupakan gambar jam kerja (X_2) dari 36 Pedagang Kaki Lima (PKL), dengan nilai rata-rata sebesar 54,31 dan standart deviasi 7,649. Dari gambar 3.2 sebaran datanya normal, karena nilai p-value > α (0,05).



Gambar 3.3 Grafik Lama Usaha PKL

Gambar 3.3 di atas merupakan gambar lama usaha (X_3) dari 36 Pedagang Kaki Lima (PKL), dengan nilai rata-rata sebesar 6,056 dan standart deviasi 1,170. Dari gambar 3.3 juga dapat dilihat bahwa sebaran tidak normal, karena p-value < α (0,05). Jika dilihat sekilas ada beberapa data yang letaknya jauh dari garis normal.

3.3.2 Deteksi *Outlier*

Untuk mengetahui pengaruh *outlier*, terlebih dahulu diduga parameter model regresi dengan menggunakan OLS yang kemudian akan dibandingkan dengan hasil pendugaan parameter model regresi dengan menggunakan metode *Least Trimmed Square* (LTS). Dari hasil analisis data didapat model regresi dengan OLS seperti pada tabel 3.1.

Tabel 3.1. Model Regresi dengan OLS

Model	R^2_{adj}	MSE
Y = - 69.6 + 0.00349 X1 + 10.2 X2 + 26.8 X3	86.6 %	1940

(3.7)

Dari model analisis regresi pengaruh antara modal PKL (X_1), jam kerja PKL (X_2), lama usaha (X_3), terhadap pendapatan bersih (Y). Dari persamaan (3.1) didapatkan β_0 sebesar -69,6, estimasi parameter β_1 sebesar 0,00349, estimasi parameter β_2 sebesar 10,2, estimasi parameter β_3 sebesar 26,8. Jika dari masing-masing X (modal PKL, jam kerja PKL, dan lama usaha PKL) akan meningkat pendapatan bersih (Y).

Dalam analisis regresi berganda untuk mendeteksi *outlier* pada peubah prediktor X yaitu dengan memeriksa nilai pengaruh (*Leverege value*) dari setiap pengamatan ke-i, kemudian membandingkannya dengan kriteria pengujian yang terdapat pada persamaan 2.11. Sedangkan untuk mendeteksi *outlier* pada peubah prediktor Y dengan memeriksa nilai mutlak ststistik uji *Studentized Deleted Residual* $|TRES|_i$ dari setiap pengamatan ke-i kemudian membandingkan dengan kriteria pengujian yang terdapat pada persamaan 2.13.

Tabel 3.2 Hasil Identifikasi *Outlier*

Pengamatan ke-	h_{ii}	$2\frac{(k+1)}{n}$	$\lvert TRES_i \rvert$	$t_{(\alpha/2;\, n-k-2)}$	Pencilan Peubah
19	0.086927	0.222	3.34517	2.04	Y, X
21	0.540345		-		X
28	0.483979		-		X

Dari hasil tabel 3.2 terdapat 3 *outlier* pada peubah prediktor X yaitu pada pengamatan 19, 21 dan 28 dan 1 *outlier* pada peubah respon Y yaitu pada pengamatan 19. Untuk itu perlu dilakukan identifikasi nilai dari data tersebut.

Sebelum mengidentifikasi *outlier* berpengaruh, maka terlebih dahulu dilakukan pendeteksian pengamatan berpengaruh. Hal ini dikarenakan *outlier berpengaruh* merupakan *outlier* yang sekaligus pengamatan berpengaruh. Pendeteksian pengamatan berpengaruh dilakukan dengan membandingkan nilai $|DFFITS|$ dari semua pengamatan ke-i pada masing-masing data dan membandingkan dengan kriteria pengujian untuk $|DFFITS|$.

Tabel 3.3 Hasil Pengamatan Berpengaruh

Pengamatanke-	$\|DFFITS_i\|$	$2\sqrt{\frac{k+1}{n}}$
19	1.03215	0.667
21	1.66671	
28	1.81726	

Dari data di atas terdapat pengamatan yang berpengaruh. Dimana pengamatan berpengaruh terhadap data. Untuk mendeteksi *outlier* dengan melihat pengamatan yang berpengaruh terhadap nilai dugaan Y dengan melihat nilai $|DFFITS|$ yang signifikan. Terdapat 3 pengamatan yang berpengaruh yaitu pengamatan 19, 21 dan 28.

Dimana dari hasil $|DFFITS|$ dari data ke 19 sebesar 1.03215 dan lebih besar dari nilai $2\sqrt{\frac{k+1}{n}}$, data ke-21 sebesar 1.66671 dan lebih besar dari nilai $2\sqrt{\frac{k+1}{n}}$, data ke-28 sebesar 1.81726 dan lebih besar dari nilai $2\sqrt{\frac{k+1}{n}}$. Sehingga data 19, 21, dan 28 termasuk dalam pengamatan yang berpengaruh karena nilai $|DFFITS|$ lebih besar dari nilai $2\sqrt{\frac{k+1}{n}}$.

3.3.3. Analisis Data dengan Metode LTS

Estimasi parameter model regresi dengan metode LTS, hanya menggunakan sub gugus pengamatan *h*. Untuk mendapatkan estimator yang *robust* terhadap pengaruh *outlier* sekaligus menghasilkan nilai *breakdown* yang optimal maka digunakan *h* sebesar $\left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right]$.

Tabel 3.4 Nilai *Breakdown* Data

n	*h*	*k*	Nilai *breakdown*
36	20	3	0.44

Penggunaan nilai *h* yang optimal didapatkan nilai *breakdown* data yang mengandung *outlier* cukup besar (tabel 3.4) dan dekat dengan nilai maksimum yaitu 0.5. Hal ini menunjukkan bahwa estimasi menggunakan metode LTS pada data yang mengandung *outlier* bersifat kekar atau *resistance* dari pada menggunakan metode OLS (tabel 3.5).

Tabel 3.5 Mode Regresi dengan OLS dan LTS

Metode	Model	R^2_{adj}	MSE
OLS	$\mathbf{Y} = -59.6 + 0.00349\, X_1 + 0.2\, X_2 + 26.8\, X_3$	0.866	1940
LTS	$\bar{\mathbf{Y}} = -32.7754 + 0.1188\, X_1 + 1.2916\, X_2 + 52.1663\, X_3$	0.9409	337.86

Dari tabel 3.5 dapat disimpulkan bahwa estimasi menggunakan metode LTS lebih bersifat *resistance* terhadap data yang mengandung *outlier* dibandingkan dengan metode OLS. Hal ini dapat dilihat dari koefisien determinasi korelasi (R^2_{adj}) pada metode LTS lebih besar dibandingkan dengan metode OLS (94.09 % $>$ 86.6 %). Sedangkan nilai *Mean Square Error* (MSE) pada metode LTS lebih kecil dibandingkan metode OLS (337.86 $<$ 1940).

Dari model regresi dengan metode LTS terlihat bahwa variabel yang paling berpengaruh terhadap pendapatan bersih PKL (Y) adalah modal PKL (X_1), dan jam kerja (X_3). Dari variabel Y (pendapatan bersih) dengan variabel X_1 (Modal PKL) diperoleh model regresi

$$Y = 630 + 0.0142\, X_1 \qquad (3.8)$$

Dari model regresi linier sederhana tersebut didapatkan nilai β_0 sebesar 630 dan β_1 sebesar 0.0142. Sedangkan untuk variabel Y (pendapatan bersih) dengan variabel X_2 (Jam kerja PKL) diperoleh model regresi

$$Y = -91.3 + 13.7X_2 \qquad (3.9)$$

Dari model regresi linier sederhana tersebut didapatkan nilai β_0 sebesar -91.3 dan β_1 sebesar 13.7. Dari 2 model regresi linier sederhana tersebut, dapat mengetehaui variabel bebas (X_1 dan X_2) yang paling berpengaruh terhadap nilai Y. Faktor yang paling berpengaruh terhadap pendapatan bersih PKL adalah modal PKL, karena setiap modal ditambah maka pendapatan bersih PKL akan meningkat. Lama jam kerja sebenarnya juga berpengaruh terhadap pendapatan bersih PKL tetapi jumlah pendapatan lebih meningkat saat modal dinaikkan.

Dari persamaan 3.8 dapat diperoh nilai *Y* (pendapatan bersih PKL) sebesar 637. Sedangkan dari persamaan 3.9 diperoleh nilai *Y* (pendapatan bersih PKL) sebesar 525. Jadi faktor yang paling berpengaruh terhadap pendapatan bersih PKL adalah modal, semakin besar modal maka pendapatan akan semakin meningkat.

3.4. Kajian Keagamaan

Pada BAB II telah disinggung bahwa regresi terdapat pada Al-Qur'an surat Al-Baqarah ayat 2-3. Penulis pada bab ini akan menghubungkan antara Al-Qur'an surat Al-Baqarah ayat 2-3 dengan konsep regresi dalam matematika. Konsep regresi dalam matematika ternyata telah terkonsep sejak zaman nabi Muhammad SAW. Hal tersebut terbukti dijelaskan dalam Al-Qur'an surat Al-Baqarah ayat 2-3, yang secara tidak langsung telah melahirkan konsep regresi.

ذَٰلِكَ ٱلْكِتَٰبُ لَا رَيْبَ ۛ فِيهِ ۛ هُدًى لِّلْمُتَّقِينَ ﴿٢﴾ ٱلَّذِينَ يُؤْمِنُونَ بِٱلْغَيْبِ
وَيُقِيمُونَ ٱلصَّلَوٰةَ وَمِمَّا رَزَقْنَٰهُمْ يُنفِقُونَ ﴿٣﴾

Artinya : " Kitab (Al Qur'an) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertaqwa. (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebagian rezeki yang kami anugerahkan kepada mereka."

Pengertian regresi dalam surat Al-Baqarah ayat 2-3 merupakan adanya hubungan antara ayat 2 dengan ayat 3, maksudnya adalah mereka yang bertakwa yaitu mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebagian rezeki yang dianugerahkan kepada mereka. Dari sini diketahui bahwa regresi dalam ayat tersebut merupakan hubungan dalam konsep yang sederhana dan dalam matematika digunakan untuk perhitungan-perhitungan dasar matematika.

Kaitan regresi pada surat ini terletak pada kalimat *"mereka yang bertaqwa"* dan *"mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebagian rezeki yang kami anugerahkan kepada mereka"*, kalimat tersebut menjelaskan adanya hubungan antara *"mereka yang bertaqwa"* dimisalkan variabel prediktor (*X*) dan *"mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezki yang kami anugerahkan kepada mereka"*dimisalkan variabel respon (*Y*), dan dalam matematika hubungan tersebut dinamakan regresi yang merupakan suatu alat ukur untuk mengukur ada atau tidaknya hubungan antara variabel prediktor *(X)* dan variabel respon *(Y)*.

Perbedaan regresi dalam surat Al-Baqarah dengan regresi dalam penelitian ini terletak pada objek yang diregresikan. Selain regresi, dalam penelitian ini juga menyinggung tentang *outlier*, yang mana dalam Al-Qur'an telah dijelaskan dalam surat Al-Kahfi ayat 49 sebagai berikut:

وَوُضِعَ ٱلْكِتَٰبُ فَتَرَى ٱلْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَٰوَيْلَتَنَا مَالِ هَٰذَا ٱلْكِتَٰبِ لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحْصَىٰهَا ۚ وَوَجَدُوا۟ مَا عَمِلُوا۟ حَاضِرًا ۗ وَلَا يَظْلِمُ رَبُّكَ أَحَدًا ﴿٤٩﴾

Artinya : "Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang 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."*

Surat Al-Kahfi ayat 49 di atas, Allah menjelaskan tentang bahwa kita dalam kehidupan tidak selalu menemui orang-orang yang benar, tetapi terdapat pula orang-orang yang bersalah. 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.

Setelah diuraikan di atas dapat diambil kesimpulan bahwa yang menjelaskan *outlier* adalah kalimat "*Dan diletakkanlah kitab, lalu kamu akan melihat orang-orang bersalah ketakutan terhadap apa yang (tertulis) di dalamnya*" 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 di antara data dalam model.

BAB IV

PENUTUP

4.1 Kesimpulan

Dari metode *Least Trimmed Square* pada pembahasan di bab III didapatkan estimasi parameter $\hat{\beta}_{LTS}$ yaitu $\hat{\beta}_{LTS} = (XX^T)^{-1}XY$ memenuhi syarat-syarat *unbias*, efisien, dan konsisten. Sehingga model regresi *robust* dengan metode *Least Trimmed Square* (LTS) yang didapat sudah dapat digunakan untuk mengatasi *outlier* pada model regresi *robust*.

Dari hasil implementasi regresi *Robust* dengan metode LTS dalam mengestimasi parameter regresi pada data faktor yang mempengaruhi pendapatan Pedagang Kaki Lima didapatakan model regresi LTS lebih efisien dibanding dengan model regresi OLS. Pada model regresi LTS nilai *Mean Square Error lebih* (MSE) yang dihasilkan lebih kecil dibandingkan pada model regresi OLS. Nilai Koefisisen Determinasi Terkoreksi pada model regresi LTS lebih besar nilainya dibandingkan dengan model regresi OLS. Sehingga metode LTS lebih *resistance* digunakan dalam data yang mengandung *outlier*.

4.2 Saran

Diharapkan untuk skripsi selanjutnya menggunakan estimator yang lain untuk mengatasi *outlier* untuk mencari estimasi parameternya, dan juga dapat menggunakan metode lain selain metode *Least Trimmed Square* untuk model regresi *robust*.

DAFTAR PUSTAKA

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

Algifari. 1997. *Analisis Regresi Teori Kasus dan Solusi*. Yogyakarta: BPFE.

Draper, N., dan Smith, H.. 1998. *Analisis Regresi Terapan*. Jakarta: PT Gramedia Pustaka Utama.

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

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

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

Isnaini, I.. 2000. Metode Permukaan Respon Untuk Analisa Pengaruh Modifikasi Sudut Pahat Gurdi Terhadap Keausan Pahat dan Kebulatan Lubang pada material Komposit Karbon. *Skripsi S1* tidak diterbitkan. Surabaya: FMIPA ITS.

Katsir, I.. 2004. *Tafsir Ibnu Katsir*. Bogor: Pustaka Imam Asy. Syafi'i.

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

Quth, S.. 2008. *Tafsir Fidzilalil Qur'an*. Jakarta: Gema Press.

Seber, G.. 2007. *Linier Regression Analysis*. New Zewland : Intersince.

Sembiring, R.K.. 1995. *Analisis Regresi*. Bandung: ITB.

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).

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

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

LAMPIRAN

LAMPIRAN 1.Data Pendapatan Pedagang Kaki Lima

No.	Y (Pendapatan bersih PKL(Rp))	X_1 (Modal PKL (Rp/Bulan))	X_2 (Jam kerja PKL(Jam/Minggu))	X_3 (Lama Usaha PKL (thn))
1.	509	500	45	5
2.	439	450	41	4
3.	714	1150	60	7
4.	454	485	41	4
5.	559	750	47	4
6.	554	800	48	5
7.	709	1100	55	5
8.	544	600	49	6
9.	734	1250	58	7
10.	519	750	45	5
11.	759	1500	67	7
12.	659	1100	58	6
13.	574	1000	51	6
14.	779	1650	60	7
15.	464	630	45	4
16.	799	1500	65	7
17.	754	1250	60	7
18.	474	500	45	4
19.	799	1850	54	7
20.	779	1650	65	7
21.	659	10000	49	6
22.	724	1400	56	7
23.	769	1750	63	8
24.	724	1650	65	7
25.	494	700	43	6
26.	774	1750	60	7
27.	744	1500	61	8
28.	709	10000	60	7
29.	729	1750	57	6
30.	669	1250	58	6
31.	509	500	45	6
32.	744	1475	59	7
33.	764	1650	61	7
34.	634	1100	58	5
35.	684	1250	54	6
36.	619	900	47	5

Sumber : Fatmawati (2008)

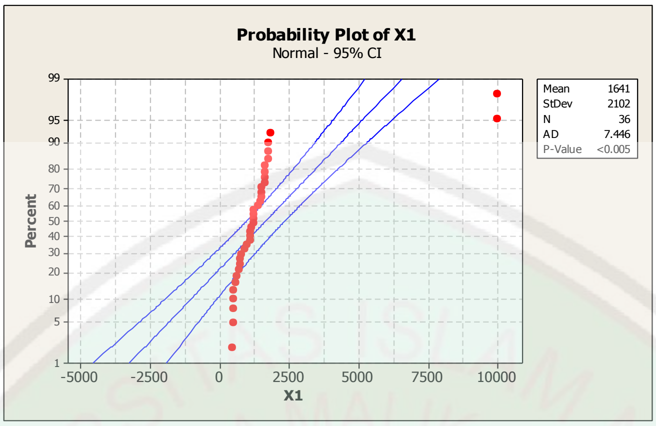
Lampiran2.Pendugaan Parameter Regresi dengan Metode OLS

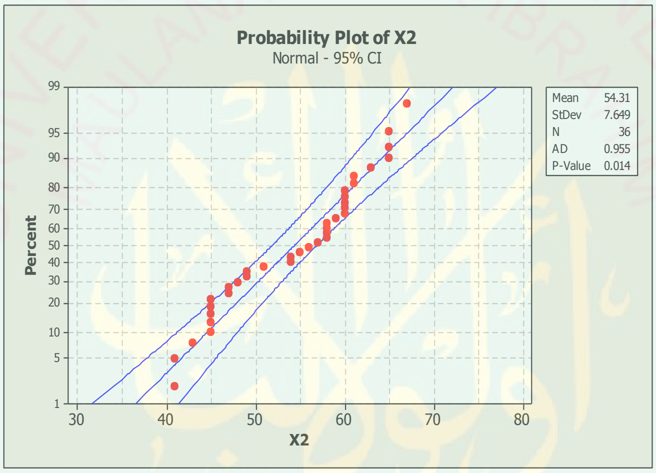
```
Welcome to Minitab, press F1 for help.
```

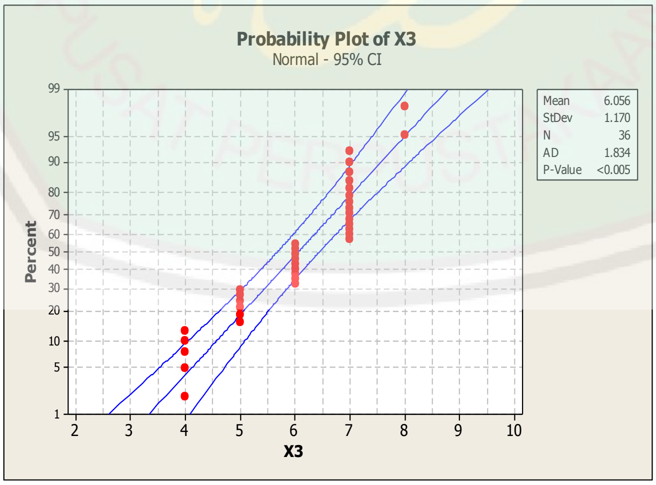
Regression Analysis: Y versus X1; X2; X3

```
The regression equation is
Y = - 69,6 + 0,00349 X1 + 10,2 X2 + 26,8 X3


Predictor       Coef    SE Coef       T      P
Constant      -69,55      53,91   -1,29  0,206
X1          0,003493   0,003669    0,95  0,348
X2            10,215      1,658    6,16  0,000
X3             26,84      11,03    2,43  0,021


S = 44,0403   R-Sq = 86,6%   R-Sq(adj) = 85,3%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       3  400573  133524  68,84  0,000
Residual Error  32   62066    1940
Total           35  462639


Source   DFSeq SS
X1        1   31237
X2        1  357843
X3        1   11493


Unusual Observations

Obs     X1       Y     Fit  SE Fit  Residual  St Resid
 19   1850  799,00  676,40   12,98    122,60     2,91R
21  10000  659,00  626,95   32,37     32,05      1,07 X
28  10000  709,00  766,16   30,64    -57,16     -1,81 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Lampiran 3. Plot antara Modal PKL (x_1), Jam Kerja (x_2) dan Lama Usaha (x_3) dengan Pendapatan (y)







Lampiran 4. Tabel Perhitungan $\hat{y}_i$, d_i, simpangan baku dari d_i ($s(d_i)$), *Studentizied Deleted Residual* (TRES), DFFITS (*The Diffrence in Fit Statistics*), dan jarak Cook.

	y	x1	x2	x3	$\hat{y}_i$	$\lvert TRES_i\rvert$	h_{ii}	$\lvert DFFITS_i\rvert$
1	509	500	45	5	525.145	0.39751	0.074246	0.11257
2	439	450	41	4	457.3705	0.46016	0.124923	0.17386
3	714	1150	60	7	734.0135	0.48979	0.052907	0.11576
4	454	485	41	4	457.49265	0.10317	0.124793	0.03896
5	559	750	47	4	519.6175	0.93685	0.134157	0.36877
6	554	800	48	5	556.792	0.08642	0.052710	0.02039
7	709	1100	55	5	629.239	1.97419	0.106855	0.68285
8	544	600	49	6	593.094	1.18890	0.071912	0.33094
9	734	1250	58	7	713.9625	0.43407	0.053852	0.10356
10	519	750	45	5	526.0175	0.18440	0.072173	0.05143
11	759	1500	67	7	806.635	1.20152	0.131525	0.46758
12	659	1100	58	6	686.639	0.66571	0.051695	0.15543
13	574	1000	51	6	614.89	0.97322	0.043806	0.20831
14	779	1650	60	7	735.7585	0.97705	0.048289	0.22008
15	464	630	45	4	498.7987	0.86000	0.119506	0.31683
16	799	1500	65	7	786.235	0.26927	0.093865	0.08666
17	754	1250	60	7	734.3625	0.42361	0.051705	0.09891
18	474	500	45	4	498.345	0.60452	0.119679	0.22289
19	799	1850	54	7	675.2565	3.34517	0.086927	1.03215
20	779	1650	65	7	786.7585	0.21296	0.093038	0.06821
21	659	10000	49	6	625.9	1.07607	0.540345	1.16671
22	724	1400	56	7	694.086	0.66930	0.065637	0.17739
23	769	1750	63	8	793.5075	0.61730	0.115001	0.22252
24	724	1650	65	7	786.7585	1.56151	0.093038	0.50013
25	494	700	43	6	532.243	0.99964	0.208100	0.51244
26	774	1750	60	7	736.1075	0.84929	0.047782	0.19025
27	744	1500	61	8	772.235	0.71463	0.133527	0.28054
28	709	10000	60	7	764.9	1.87646	0.483979	1.81726
29	729	1750	57	6	678.7075	1.14482	0.040726	0.23589
30	669	1250	58	6	687.1625	0.44476	0.050933	0.10303

31	509	500	45	6	551.945	1.08716	0.154165	0.46413
32	744	1475	59	7	724.94775	0.40984	0.049108	0.09314
33	764	1650	61	7	745.9585	0.38629	0.051571	0.09008
34	634	1100	58	5	659.839	0.66735	0.172958	0.30518
35	684	1250	54	6	646.3625	0.83785	0.028771	0.14421
36	619	900	47	5	546.941	1.71080	0.055795	0.41588

Lampiran 5. Perhitungan Nilai *h* dan Nilai *Breakdown*

1. **Mencari nilai *h***

$$h = \left[\frac{n}{2}\right] + \left[\frac{(k+1)}{2}\right],$$

dimana :

$n = 36$

$k = 3$, sehingga

$$h = \left[\frac{36}{2}\right] + \left[\frac{(3+1)}{2}\right]$$

$$= 18 + \left[\frac{4}{2}\right]$$

$$= 18 + 2$$

$$= 20$$

2. **Mencari nilai *Breakdown***

$$\varepsilon_n^* = \frac{n-h}{n}$$

$$= \frac{30-20}{36}$$

$$= \frac{16}{36}$$

$$= 0.44$$

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS & TEKNOLOGI
Jl. Gajahyana No. 50 Malang Telp. (0341) 551354 Fax. (0341) 572533

BUKTI KONSULTASI SKRIPSI

Nama : Dewi Ratnasari
NIM : 08610078
Fakultas/ Jurusan : Sains Dan Teknologi/Matematika
Judul Skripsi : Deteksi *Outlier* Pada Model Regresi *Robust* Dengan Metode *Least Trimmed Square* (LTS)
Pembimbing I : Dr. Sri Harini, M.Si
Pembimbing II : Dr. H. Ahmad Barizi, M.A

No	Tanggal	HAL	Tanda Tangan	
1	4 April 2013	Bab I	1.	
2	27 Mei 2013	Revisi Bab I dan Bab II		2.
3	3 Juni 2013	Bab I Agama	3.	
4	22 Agustus 2013	Revisi Bab II dan Bab III Agama		4.
5	27 Agustus 2013	Revisi dan konsultasi Bab III	5.	
6	18 Desember 2013	Revisi Bab III		6.
7	30 Februari 2014	Revisi dan konsultasi Bab III	7.	
8	10 Maret 2014	Revisi Bab III Agama		8.
9	11 Maret 2014	Revisi Bab III dan konsultasi Bab IV	9.	
10	12 Maret 2014	ACC BAB I, II, III, IV		10.

Malang, 07 April 2014
Mengetahui,
Ketua Jurusan Matematika

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