

**DETEKSI *OUTLIER* PADA MODEL REGRESI TERBOBOTI
DENGAN METODE *RIDGE***

SKRIPSI

Oleh:
KURNIA IRIANTI
NIM. 08610026



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

**DETEKSI *OUTLIER* PADA MODEL REGRESI TERBOBOTI
DENGAN METODE *RIDGE***

SKRIPSI

Diajukan kepada:
Univeritas Islam Negeri Maulana Malik Ibrahim Malang
untuk memenuhi salah satu persyaratan
dalam memperoleh gelar Sarjana Sains (S.Si)

Oleh:
KURNIA IRIANTI
NIM. 08610026

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

**DETEKSI *OUTLIER* PADA MODEL REGRESI TERBOBOTI
DENGAN METODE *RIDGE***

SKRIPSI

oleh:
KURNIA IRIANTI
NIM. 08610026

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 13 Agustus 2012

Pembimbing I

Pembimbing II

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

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

Mengetahui,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP. 19751006 200312 1 001

DETEKSI *OUTLIER* PADA MODEL REGRESI TERBOBOTI DENGAN METODE *RIDGE*

SKRIPSI

Oleh:
KURNIA IRIANTI
NIM. 08610026

Skripsi ini telah dipertahankan di depan Dewan Penguji
dan dinyatakan diterima sebagai salah satu persyaratan
dalam memperoleh gelar Sarjana Sains (S.Si)
Tanggal: 8 September 2012

Susunan Dewan Penguji

TandaTangan

- | | | |
|------------------|--|-------|
| 1. Penguji Utama | : <u>Abdul Aziz, M.Si</u>
NIP. 19760318 200604 1 002 | _____ |
| 2. Ketua | : <u>Drs. H. Turmudi, M.Si</u>
NIP. 19571005 198203 1 006 | _____ |
| 3. Sekretaris | : <u>Dr. Sri Harini, M.Si</u>
NIP. 19731014 200112 2 002 | _____ |
| 4. Anggota | : <u>Dr. H. Munirul Abidin, M.Ag</u>
NIP. 19720420 200212 1 003 | _____ |

Mengetahui,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP. 19751006 200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Kurnia Irianti

NIM : 08610026

Jurusan : Matematika

Fakultas : Sains dan Teknologi

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, 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, 15 Agustus 2012

Yang membuat pernyataan,

Kurnia Irianti
NIM. 08610026

MOTTO

- ***Kegagalan hanya terjadi bila kita menyerah*** -



HALAMAN PERSEMBAHAN

Skripsi ini dipersembahkan untuk:

Kedua orang tua penulis, permata tercinta Ayahanda H. Mohamad Irfan, S.H dan Ibunda Hj. Ninik Purbowiningsih, S.H yang tidak pernah berhenti memberikan kasih sayang, do'a, dan semangat kepada penulis baik moril maupun materiil.



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. Imam Suprayogo, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang telah banyak memberikan pengetahuan dan pengalaman yang berharga.
2. Prof. Drs. Sutiman Bambang Sumitro, SU., D.Sc, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Abdussakir, M.Pd, selaku Ketua Jurusan Matematika yang telah memberikan pengarahan dan pengalaman yang berharga.
4. Dr. Sri Harini, M.Si dan Dr. H. Munirul Abidin, M.Ag, selaku dosen pembimbing skripsi, yang telah memberikan banyak pengarahan dan pengalaman yang berharga.

5. Abdul Azis, M.Si dan Drs. H. Turmudzi, M.Si, selaku tim penguji skripsi, terimakasih telah memberikan masukan-masukan yang berharga dan bermanfaat untuk penulisan skripsi ini.
6. Seluruh dosen jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah membantu dalam menyelesaikan skripsi ini.
7. H. Mohamad Irfan, S.H dan Hj. Ninik Purbowiningsih, S.H, yang telah mencurahkan cinta dan kasih sayang teriring do'a, motivasi, dan materi, sehingga penulis selalu optimis dalam menggapai salah satu kesuksesan hidup.
8. Kakak dan Adik penulis, Annisa Kurnia Indah, S.Farm, Dita Amalia dan Aulia Sandra Ozha yang telah memberikan dukungan, do'a, motivasi dan materi bagi penulis.
9. Teman-teman terbaik penulis, Shofiatul Inayah, Elva Ravita Sari, Amilatul Zakiyah, Anang Fakhmi, Oky Dwi Ardian dan Tri Wahyudianto, serta seluruh teman-teman jurusan matematika khususnya angkatan 2008 yang berjuang bersama-sama untuk mencapai kesuksesan yang diimpikan. Terimakasih atas segala pengalaman berharga dan kenangan terindah yang telah terukir.
10. Seluruh penghuni Kost Ampel I/1 yang telah menjadi penyemangat dan penghibur lika-liku kehidupan penulis.
11. Kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, yang tidak bisa disebutkan satu per satu.

Akhirnya dengan segala keterbatasan pengetahuan dan waktu penulis, sekiranya ada sesuatu yang kurang berkenan sehubungan dengan penyelesaian skripsi ini, penulis mohon maaf yang sebesar-besarnya. Kritik dan saran dari para pembaca yang budiman demi kebaikan karya ini merupakan harapan besar bagi penulis. Semoga karya ilmiah yang berbentuk skripsi ini dapat bermanfaat dan berguna.

Malang, 15 Agustus 2012

Penulis,



DAFTAR ISI

HALAMAN JUDUL	
HALAMAN PENGAJUAN	
HALAMAN PERSETUJUAN	
HALAMAN PENGESAHAN	
HALAMAN PERNYATAAN KEASLIAN TULISAN	
MOTTO	
HALAMAN PERSEMBAHAN	
KATA PENGANTAR	viii
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
ABSTRAK	xv
ABSTRACT	xvi
الملخص	xvii
 BAB I : PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan masalah	5
1.3. Tujuan Penelitian	5
1.4. Batasan Masalah	5
1.5. Manfaat Penelitian	5
1.6. Metode Penelitian	6
1.7. Sistematika Penulisan	7
 BAB II : KAJIAN PUSTAKA	
2.1. Regresi Linier	9
2.2. Metode Kuadrat Terkecil Terboboti.....	10
2.3. <i>Outlier</i>	13
2.4. Regresi <i>Ridge</i>	15
2.5. Kajian Masalah Regresi dan <i>Outlier</i> dalam Al-Qur'an.....	17
2.5.1 Analisis Regresi	17

2.5.2 Kajian Keislaman Tentang <i>Outlier</i>	19
---	----

BAB III : PEMBAHASAN

3.1 Menentukan <i>Outlier</i> pada Model Regresi Linier Terboboti dengan Metode <i>Ridge</i>	22
3.2 Menentukan Estimasi Parameter Model Regresi Linier Terboboti yang mengandung <i>Outlier</i>	24
3.3 Menentukan Sifat-Sifat Estimasi Parameter Regresi Linier yang mengandung <i>Outlier</i>	26
3.4 Aplikasi pada Estimasi Parameter Model Regresi Linier yang mengandung <i>Outlier</i>	29
3.5 Keterkaitan Hasil Penelitian Dengan Kajian Agama	39

BAB IV : PENUTUP

4.1 Kesimpulan	41
4.2 Saran	41

DAFTAR PUSTAKA

BUKTI KONSULTASI SKRIPSI

DAFTAR GAMBAR

Gambar 3.1 Scatterplot of X1	33
Gambar 3.2 Scatterplot of X2	33
Gambar 3.3 Scatterplot of X3	34
Gambar 3.4 Scatterplot of X4	34
Gambar 3.5 Scatterplot of Y	35
Gambar 3.6 Boxplot of X1, X2.....	36
Gambar 3.7 Boxplot of X3, X4.....	36
Gambar 3.8 Boxplot of Y.....	37
Gambar 3.9 Probability Plot of Y	38
Gambar 3.10 Probability Plot of X1	38
Gambar 3.11 Probability Plot of X2	39
Gambar 3.12 Probability Plot of X3	39
Gambar 3.13 Probability Plot of X4	40
Gambar 3.14 Hasil Analisis Regresi Linier dengan Minitab	40

DAFTAR TABEL

Tabel 3.1 Data Tes Fungsi Enzim	31
---------------------------------------	----



ABSTRAK

Irianti, Kurnia. 2012. **Deteksi *Outlier* pada Model Regresi Terboboti dengan Metode *Ridge***. Skripsi. Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (I) Dr. Sri Harini, M.Si
(II) Dr.H.Munirul Abidin, M.Ag

Kata Kunci: *Outlier, ridge regression*, regresi terboboti

Model regresi terboboti adalah suatu model untuk memboboti yang dapat ditentukan berdasarkan data pengamatan. Pada penelitian ini model regresi terboboti yang digunakan untuk mendeteksi *outlier* adalah

$$\tilde{y} = \mathbf{X}\mathbf{W}^T\tilde{\beta} + \tilde{\varepsilon}$$

dimana dari model tersebut akan didekati dengan metode *ridge*.

Metode *Ridge* merupakan modifikasi dari metode kuadrat terkecil dengan cara menambah tetapan bias c pada diagonal matrik $\mathbf{X}^T\mathbf{X}$. Sehingga sifat dari model tersebut menjadi robust.

Dari hasil penelitian didapatkan estimasi parameter model regresi terboboti dengan metode *ridge* adalah :

$$\tilde{\beta}^* = (\mathbf{X}^T\mathbf{W}\mathbf{X} + \tilde{\beta}_{WLS}\mathbf{I}_r)^{-1}(\mathbf{X}^T\mathbf{W})\tilde{y}$$

yang memenuhi syarat *unbias*. Sehingga $\tilde{\beta}^*$ yang didapat sudah dapat digunakan untuk mengatasi *outlier* pada model regresi terboboti.

ABSTRACT

Irianti, Kurnia. 2012. Outlier Detection on Weighted Regression Models with Ridge Method. Thesis. Department of Mathematics. Faculty of Science and Technology. State Islamic University Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Sri Harini, M.Si
(II) Dr.H.Munirul Abidin, M.Ag

Keywords: Outlier, ridge regression, weighted regression

Weighted regression model is a model for weighing that can be determined based on observational data. In this study, weighted regression models used to detect outliers is

$$\tilde{y} = \mathbf{XW}^T\tilde{\beta} + \tilde{\varepsilon}$$

where the model will be approximated by ridge method.

Ridge method is a modification of the method of least squares method by adding a constant bias c on the matrix diagonal $\mathbf{X}^T\mathbf{X}$, thus the nature of the model becomes robust.

In the research finding, it is found that estimated parameters of the regression model by weighted ridge regression method, namely:

$$\tilde{\beta}^* = (\mathbf{X}^T\mathbf{W}\mathbf{X} + \tilde{\beta}_{WLS}\mathbf{I}_r)^{-1}(\mathbf{X}^T\mathbf{W})\tilde{y}$$

these results obtained parameter estimates that are *unbias* qualify. Thus, the obtained one can be used to handle the outliers on weighted regression models.

الملخص

ارينتي،كورني. ٢٠١٢. كشف الخارجية على نماذج الانحدار الموزون مع ريدج الطريقة. البحث الجامعي. قسم الرياضيات لكلية العلوم والتكنولوجيا التابعة لجامعة ولاية الإسلامية مولانا مالك ابراهيم مالانج. المشرف: (١) دبسري هارينى الماجستير (٢) د.ج. مونرؤل ابدن الماجستير

الكلمة الرئيسية : أوتلير، ريدج الانحدار، الانحدار الموزون

المرجح نموذج الانحدار هو نموذج لوزن والذي يمكن تحديده على أساس بيانات الرصد. في هذا البحث، تم استخدام نموذج الانحدار المرحج للكشف عن القيم المتطرفة و

$$\tilde{y} = \mathbf{XW}^T\tilde{\beta} + \tilde{\varepsilon}$$

وسيتم تناول أي من هذه النماذج من طريقة ريدج.

$\mathbf{X}^T\mathbf{X}$ ريدج هو تعديل لطريقة المربعات الصغرى بإضافة ج التحيز المستمر على مصفوفة قطري. وبالتالي فإن طبيعة النموذج لتكون قوية.

وأظهرت نتيجة معلمة نموذج الانحدار موازنة تقديرات طريقة ريدج هو:

$$\tilde{\beta}^* = (\mathbf{X}^T\mathbf{W}\mathbf{X} + \tilde{\beta}_{WLS}\mathbf{I}_r)^{-1}(\mathbf{X}^T\mathbf{W})\tilde{y}$$

اونبيس المؤهلة. يمكن الحصول بالفعل وبالتالي يمكن استخدامها للتعامل مع القيم المتطرفة على نماذج الانحدار المرحج.

BAB I PENDAHULUAN

1.1. Latar Belakang

Matematika adalah suatu ilmu pasti atau eksak yang mempunyai rumus-rumus dalam penyelesaiannya. Kata matematika berasal dari bahasa Yunani yaitu *mathematika* yang berarti studi besaran, ruang, struktur dan perubahan. Pada masa kini ilmu matematika telah banyak mengalami kemajuan. Sebagai ilmu yang dikembangkan manusia, matematika secara tersurat dan tersirat ada di dalam Al-Qur'an.

Sebagaimana Allah SWT berfirman dalam QS. An-Nahl : 125 sebagai berikut.

أَدْعُ إِلَى سَبِيلِ رَبِّكَ بِالْحُكْمَةِ وَالْمَوْعِظَةِ الْحَسَنَةِ ۚ وَجِدِلْهُمْ بِالَّتِي هِيَ
أَحْسَنُ ۚ إِنَّ رَبَّكَ هُوَ أَعْلَمُ بِمَن ضَلَّ عَنْ سَبِيلِهِ ۚ وَهُوَ أَعْلَمُ بِالْمُهْتَدِينَ ﴿١٢٥﴾

Artinya : *“Serulah (manusia) kepada jalan Tuhan-mu dengan hikmah dan pelajaran yang baik dan bantahlah mereka dengan cara yang baik. Sesungguhnya Tuhanmu Dialah yang lebih mengetahui tentang siapa yang tersesat dari jalan-Nya dan Dialah yang lebih mengetahui orang-orang yang mendapat petunjuk.”*

Ayat ini berlaku umum untuk sasaran dakwah siapa saja, muslim ataupun kafir. Meski ayat ini adalah perintah Allah SWT kepada Rasulullah, perintah ini juga berlaku untuk umat Islam. Secara umum Allah memerintahkan Rasul-Nya Muhammad untuk menyeru manusia ke Jalan Allah dengan hikmah.

Hal tersebut juga telah dikisahkan Rasulullah untuk selalu berbuat amar makruf nahi munkar seperti dalam hadits berikut ini :

وَالَّذِي نَفْسِي بِيَدِهِ لَتَأْمُرَنَّ بِالْمَعْرُوفِ وَلَتَنْهَوْنَ عَنِ الْمُنْكَرِ أَوْ لَيُوشِكَنَّ اللَّهُ أَنْ يَبْعَثَ عَلَيْكُمْ عِقَابًا مِنْهُ ثُمَّ تَدْعُوهُ فَلَا يُسْتَجَابُ لَكُمْ

Artinya : *“Demi Dzat yang diriku berada di tangan-Nya, sungguh kalian (mempunyai dua pilihan, yaitu) melakukan amar ma’ruf nahi munkar ataukah Allah mendatangkan kepada kalian siksa dari-Nya yang menimpa kalian. Kemudian setelah itu kalian berdoa, maka (doa itu) tidak akan dikabulkan”* (HR Ahmad dan al-Tirmidzi).

Hadits ini menjadi qarinah aktivitas amar ma’ruf nahi munkar berhukum fardhu. Karena aktivitasnya fardhu, maka keberadaan jamaah yang mengerjakan aktivitas itu juga wajib.

Untuk menyeru kepada kebenaran Islam dan kepada jalan Allah mestilah dengan cara yang benar seperti yang pernah dicontohkan oleh Rasulullah ketika melaksanakan perintah ini. Jika kita berharap bahwa dakwah kita bernilai ibadah, maka mestilah dakwah itu tidak menyimpang dari contoh pelaksanaan Nabi dan para sahabat. Begitu pula dalam menghadapi bantahan atau sanggahan dari *ahlul kitab*, mestilah dengan yang lebih baik kecuali orang yang *dhalim* di antara mereka.

Dakwah adalah kumpulan ilmu pengetahuan yang dikembangkan umat Islam secara sistematis metodologis. Ilmu pengetahuan yang dimaksudkan adalah ilmu-ilmu eksak, seperti ilmu statistika, matematika, arsitektur, informatika, fisika, biologi, kimia, ilmu falak, kedokteran, maupun ilmu-ilmu sosial, seperti sosiologi, psikologi, ekonomi, sastra, dan lain-lain. Termasuk pula di dalamnya adalah ilmu-ilmu agama, seperti

aqidah, ibadah, akhlaq, muamalah, fiqh, dan ilmu-ilmu lainnya yang terkandung di dalam Al-Quran.

Al-Quran merupakan mukjizat terbesar yang diberikan kepada Nabi Muhammad SAW. Di antara kemukjizatan Al-Qur'an adalah kebenaran ayat-ayatnya yang kemudian terungkap satu persatu sejalan dengan ilmu pengetahuan modern. Mungkin dalam suatu penggalan sejarah tertentu sains tidak mampu mengungkap kebenaran ini. Di belakang hari baru terbukti, dan menjadi jelaslah bagi manusia bahwa apa yang diberitakan Al-Qur'an adalah benar.

Kaitan ayat Al-Qur'an tersebut dengan ilmu pengetahuan yaitu sesungguhnya manusia diperintahkan untuk menyeru pada agama Allah, dan ajakan kepada orang-orang yang menyimpang untuk kembali ke jalan Allah yang lurus. Hal ini sejalan dengan kondisi suatu penelitian, dimana banyak penyimpangan data (*outlier*) yang terjadi dan membutuhkan penyelesaian. Hal tersebut dapat terjadi misalnya karena kesalahan dalam pencatatan pengamatannya, kesalahan alat ukur, atau karena ketidakcocokan model yang digunakan.

Untuk mendeteksi suatu data mengandung *outlier* atau tidak, selama ini digunakan model regresi terboboti. *Outlier* dalam model regresi terboboti dapat berpengaruh pada hasil analisis statistik. Jadi model Regresi terboboti adalah suatu model untuk memboboti (w_i) yang dapat ditentukan berdasarkan data pengamatan. Akan tetapi terkadang data *outlier* yang sudah diboboti tidak menunjukkan adanya perubahan. Hal ini dapat disebabkan karena pada model tersebut masih terdapat bias pada *error*.

Sehingga penggunaan model regresi terboboti membutuhkan penyelesaian dengan menggunakan model pendeteksi matematis yang lebih akurat dan tepat.

Ada banyak cara untuk mengatasi hal ini, di antaranya ialah dengan memperbesar ukuran sampel sehingga kovarian diantara parameter-parameter-nya dapat dikurangi, tetapi hal ini akan benar jika interkorelasi yang terjadi hanya didalam sampel dan bukan didalam populasi dari variabel-variabel. Jika variabel-variabel ini berkolinier dalam populasi maka prosedur memperbesar ukuran sampel tidak akan mengurangi *outlier*. Atau dengan menggunakan metode Regresi *Ridge* (Hoerl dan Kennard, 1970). Regresi ini merupakan modifikasi dari metode kuadrat terkecil dengan cara menambah tetapan bias c yang kecil pada diagonal matriks $X^T X$. Sifat dari metode ini adalah robust terhadap *outlier*.

Dari beberapa cara mengatasi masalah *outlier* di atas, metode Regresi *Ridge* merupakan penyelesaian yang paling baik, karena mengingat tujuan Regresi *Ridge* adalah untuk memperkecil variansi estimator koefisien regresi. Dengan menggunakan metode Regresi *Ridge* maka estimasi koefisien regresi yang dihasilkan lebih stabil dan lebih tepat (Hoerl dan Kennard, 1970)

Berdasarkan uraian di atas, sehingga pada tugas akhir ini penulis tertarik untuk mengambil judul penelitian **“Deteksi *Outlier* pada Model Regresi Terboboti dengan Metode *Ridge*.”**

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka dalam penelitian ini rumusan masalah yang akan dibahas adalah bagaimana estimasi parameter model regresi terboboti yang mengandung *outlier* dengan metode *Ridge*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan yang ingin dicapai dalam penelitian ini adalah mendapatkan estimasi parameter model regresi terboboti yang mengandung *outlier* dengan metode *Ridge*.

1.4. Batasan Masalah

Sesuai rumusan masalah dan tujuan penelitian, serta agar pembahasan lebih fokus maka pembatasan masalah yang diberikan adalah :

1. Model regresi yang digunakan pada penelitian ini adalah model regresi linier
2. *Error* diasumsikan berdistribusi normal dengan mean μ dan varian σ^2

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Bagi penulis

Penulis mengetahui tentang pendeteksian *outlier* pada model regresi terboboti dengan metode *ridge*. Dapat menjadi wacana baru dalam pengembangan ilmu pengetahuan khususnya ilmu matematika yang dapat dimanfaatkan dalam kehidupan sehari-hari.

2. Bagi lembaga

Sebagai sumbangan pemikiran dan sebagai upaya peningkatan kualitas keilmuan khususnya dalam bidang matematika di jurusan Matematika Fakultas Sains dan Teknologi UIN MALIKI Malang.

3. Bagi pembaca

Memberikan gambaran tentang pendeteksian *outlier* pada model regresi terboboti dengan metode *Ridge*, sehingga pembaca dapat mengaplikasikan pada bidang yang lain dan menggunakan distribusi yang lain bila ingin mengembangkan metode *Ridge*.

1.6. Metode Penelitian

1.6.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan penelitian kepustakaan. Dimana penelitian kepustakaan ini merujuk pada pustaka atau buku-buku yang berkaitan dan yang dibutuhkan untuk melakukan penelitian ini. Untuk mendeteksi *outlier* pada model regresi terboboti dengan metode *Ridge*, terlebih dahulu dikaji mengenai definisi dan sifat dasar dari analisis regresi terboboti, dan metode *Ridge*. Selanjutnya dianalisis dan bila perlu dilakukan studi simulasi untuk memberikan gambaran yang lebih jelas penggunaan metode dan sifat-sifat dari metode *Ridge*.

1.6.2. Langkah Analisis

Metode penelitian yang digunakan dalam penelitian ini adalah metode kepustakaan yaitu usaha mendalami, mencermati, menelaah dan mengidentifikasi pengetahuan yang ada dalam

kepastakaan (sumber bacaan, buku-buku referensi atau hasil penelitian dari orang lain) sebagai literatur untuk mengumpulkan data-data dan informasi (Hasan, 2002: 45).

Adapun langkah-langkah dalam penelitian ini adalah:

1. Menentukan model regresi terboboti yang mengandung *outlier*.
2. Mengestimasi model persamaan regresi terboboti menggunakan metode *Ridge*.
3. Menerapkan model yang telah dianalisis ke dalam data yang mengandung *outlier*.
4. Menganalisis model persamaan regresi terboboti menggunakan metode *Ridge*.
5. Membuat kesimpulan yang merupakan jawaban dari permasalahan yang telah dikemukakan pada pembahasan.

1.7. Sistematika Penulisan

Untuk mempermudah memahami penulisan ini secara keseluruhan, maka penulis menggambarkan sistematika penulisannya sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini menyajikan tentang teori-teori mengenai pendeteksian outlier pada model regresi terboboti dengan metode *Ridge* yang diambil dari beberapa referensi yang terkait dengan topik tersebut.

BAB III PEMBAHASAN

Pada bab ini membahas tentang pendeteksian outlier pada model regresi terboboti dengan metode *Ridge*.

BAB IV PENUTUP

Pada bab ini berisi kesimpulan dari pembahasan dan saran-saran yang berkaitan dengan penelitian ini.



BAB II

KAJIAN PUSTAKA

1.1. Regresi Linier

Analisis regresi merupakan suatu alat ukur untuk mengukur ada atau tidaknya hubungan antara variabel prediktor (X) dan variabel respon (Y). Istilah regresi yang berarti ramalan atau taksiran pertama kali diperkenalkan oleh Sir Francis Galton dalam artikelnya "*Family Likeness in Stature*" pada tahun 1886. Tujuan dari analisis regresi ini adalah untuk membuat model terbaik sehingga dapat menggambarkan bentuk antarvariabel. Dengan mengetahui adanya hubungan antarvariabel, maka dapat dilakukan estimasi suatu variabel berdasarkan variabel lain melalui persamaan yang dibuat atas hubungan tersebut. (Algifari, 1997)

Secara umum model regresi linier dapat dinyatakan sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.1)$$

dimana :

Y = variabel respon

X = variabel prediktor

β = parameter model

ε = error

Menurut Draper dan Smith (1992:189) persamaan regresi linear sederhana (2.1) dapat diubah dalam bentuk matriks yaitu :

$$\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon} \quad (2.2)$$

Atau dapat disusun sebagai berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{k1} \\ 1 & X_{21} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} x \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.3)$$

dengan :

$\tilde{y}$ = vektor variabel respon berdimensi $n \times 1$

$\mathbf{X}$ = matriks variabel prediktor berdimensi $n \times q$, dimana $q = k + 1$

$\tilde{\beta}$ = variabel parameter regresi berdimensi $q \times 1$

$\tilde{\varepsilon}$ = error

1.2. Metode Kuadrat Terkecil Terboboti

Metode kuadrat terkecil terboboti adalah suatu metode untuk memboboti (w_i) yang dapat ditentukan berdasarkan data pengamatan. Draper dan Smith (1992:103) menyatakan bahwa pembobot diberikan agar ditemukan model baru yang memenuhi asumsi dari model tersebut. Sehingga pada model tersebut dapat diterapkan hal-hal yang bersangkutan dengan metode kuadrat terkecil.

Secara matematis persamaan model regresi linier terboboti dapat dinyatakan sebagai berikut :

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 W_1 + \cdots + \beta_k X_k W_k + \varepsilon \\ Y_1 &= \beta_0 + \beta_1 X_{11} W_{11} + \cdots + \beta_k X_{k1} W_{k1} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{12} W_{12} + \cdots + \beta_k X_{k2} W_{k2} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{1n} W_{1n} + \cdots + \beta_k X_{kn} W_{kn} + \varepsilon_n \end{aligned} \quad (2.4)$$

dengan misalkan dalam model terdapat n pengamatan, dinyatakan dalam bentuk matriks sebagai berikut :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11}W_{11} & X_{21}W_{21} & \dots & X_{k1}W_{k1} \\ 1 & X_{12}W_{12} & X_{22}W_{22} & \dots & X_{k2}W_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n}W_{1n} & X_{2n}W_{2n} & \dots & X_{kn}W_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k1} & W_{k2} & \dots & W_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2.5)$$

Persamaan (2.5) dapat disederhanakan menjadi

$$\tilde{y} = \mathbf{XW}^T \tilde{\beta} + \tilde{\varepsilon} \quad (2.6)$$

yang diasumsikan $\tilde{\varepsilon} \sim N(0, \sigma^2)$ dan

$$\begin{aligned} E(\tilde{y}) &= E(\mathbf{XW}^T \tilde{\beta} + \tilde{\varepsilon}) \\ &= E(\mathbf{XW}^T \tilde{\beta}) + E(\tilde{\varepsilon}) \\ &= \mathbf{XW}^T \tilde{\beta} \end{aligned} \quad (2.7)$$

dimana varian dari *error* dapat dinyatakan sebagai berikut :

$$Var(\varepsilon) = \begin{bmatrix} var(\varepsilon_1) & cov(\varepsilon_1, \varepsilon_2) & \dots & cov(\varepsilon_1, \varepsilon_n) \\ cov(\varepsilon_2, \varepsilon_1) & var(\varepsilon_2) & \dots & cov(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\varepsilon_n, \varepsilon_1) & cov(\varepsilon_n, \varepsilon_2) & \dots & var(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.8)$$

Untuk mendapatkan estimasi parameter $\tilde{\beta}$ dari model Regresi Terboboti, maka model (2.7) dapat dicari nilai kuadrat *error*nya dengan cara sebagai berikut :

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = (\tilde{y} - \mathbf{XW} \tilde{\beta})^T (\tilde{y} - \mathbf{XW} \tilde{\beta}) \quad (2.9)$$

dengan

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = S \quad (2.10)$$

Persamaan (2.9) adalah skalar, sehingga komponen-komponennya juga skalar. Akibatnya, transpose skalar tidak merubah nilai skalar tersebut.

Sehingga S dapat ditulis sebagai berikut :

$$\begin{aligned}
S &= (\tilde{y} - \mathbf{XW}\tilde{\beta})^T (\tilde{y} - \mathbf{XW}\tilde{\beta}) \\
&= \tilde{y}^T \tilde{y} - 2\tilde{y}^T \mathbf{XW}\tilde{\beta} + \tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta}
\end{aligned} \tag{2.11}$$

Untuk meminimumkannya dapat diperoleh dengan melakukan turunan pertama S terhadap $\tilde{\beta}$,

$$\begin{aligned}
\frac{dS}{d\tilde{\beta}} &= 0 - 2\tilde{y}^T \mathbf{XW} + \mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta} \\
&= -2\tilde{y}^T \mathbf{XW} + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta}
\end{aligned} \tag{2.12}$$

dan menyamakannya dengan nol diperoleh

$$\begin{aligned}
\mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta} &= \tilde{y}^T \mathbf{XW} \\
\mathbf{X}^T \mathbf{W}\tilde{\beta} &= \mathbf{X}^T \mathbf{W}\tilde{y}
\end{aligned} \tag{2.13}$$

yang dinamakan sebagai persamaan normal, dan

$$\tilde{\beta}_{WLS} = (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W})\tilde{y} \tag{2.14}$$

penaksir parameter $\tilde{\beta}$ pada model (2.14) merupakan penaksir $\tilde{\beta}$ dari model regresi terboboti.

1.3. *Outlier*

Outlier (pencilan) didefinisikan sebagai suatu pengamatan yang tampak bertentangan atau tidak konsisten terhadap pengamatan yang lain. 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 univariat atau multivariat. *Outlier* pada model linier multivariat dapat dibagi atas tiga kategori, yaitu *outlier* terhadap leverage dan *error* ataupun keduanya (Soemarti, 2007).

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

Sebagaimana dikemukakan oleh 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.

Outlier dapat diartikan data yang tidak mengikuti pola umum pada model atau data yang keluar dari model dan tidak berada dalam daerah selang kepercayaan (Sembiring, 1995:62). Oleh karenanya, suatu *outlier* patut diperiksa secara seksama, barangkali saja alasan di balik keganjilan itu dapat diketahui (Draper dan Smith, 1992:146).

Berbagai kaidah telah diajukan untuk menolak *outlier* (dengan kata lain untuk memutuskan menyisihkan amatan tersebut dari data, untuk kemudian menganalisis kembali tanpa amatan tersebut). Penolakan begitu saja suatu pencilan bukanlah prosedur yang bijaksana. Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik data lainnya, misalnya karena pencilan timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. Sebagai kaidah umum, pencilan baru kita tolak jika setelah ditelusuri ternyata

merupakan akibat dari kesalahan-kesalahan seperti kesalahan mencatat amatan bersangkutan atau kesalahan ketika menyiapkan peralatan. Bila ternyata bukan akibat dari kesalahan-kesalahan semacam itu, penyelidikan yang saksama harus dilakukan (Draper dan Smith, 1992:146)

1.4. Regresi *Ridge*

Regresi *Ridge* adalah suatu teknik yang dikembangkan untuk menstabilkan koefisien regresi karena adanya multikolinearitas dengan cara memodifikasi metode kuadrat terkecil, sehingga dihasilkan penduga koefisien regresi lain yang bias.

Metode Regresi *Ridge* pertama kali dikemukakan oleh A.E. Hoerl pada tahun 1962. Metode ini ditujukan untuk mengatasi kondisi buruk (*ill-conditioned*) yang diakibatkan oleh korelasi yang tinggi antara beberapa peubah bebas di dalam model regresi, sehingga menyebabkan hasil matriks $\mathbf{X}^T\mathbf{X}$ -nya hampir singular, yang pada gilirannya menghasilkan nilai dugaan parameter model regresi yang tidak stabil. Umumnya sifat dari penafsiran ridge ini memiliki variansi yang minimum (Draper dan Smith, 1992:299).

Regresi *Ridge* merupakan modifikasi dari metode kuadrat terkecil yang menghasilkan penduga bias dari koefisien regresi. Modifikasi tersebut ditempuh dengan cara menambah tetapan bias yang relatif kecil pada diagonal utama matriks $\mathbf{X}^T\mathbf{X}$, sehingga koefisien estimator Ridge dipenuhi dengan besarnya tetapan bias tersebut.

Dengan membentuk $\mathbf{X}^T\mathbf{X}$ menjadi bentuk matriks korelasi, maka kesalahan yang disebabkan pengaruh pembulatan menjadi lebih kecil (Draper dan Smith, 1992:299). Terutama jika variabel *regressornya* lebih

dari dua dan data yang ada besar. Jika $\mathbf{X}^T\mathbf{X}$ yang merupakan matriks korelasi adalah matriks identitas maka nilai dengan variabel *regressand* akan sama dengan nilai sebenarnya.

Apabila $\mathbf{X}^T\mathbf{X}$ tidak mendekati matriks identitas melainkan menjauhinya, maka dapat dikatakan $\mathbf{X}^T\mathbf{X}$ hampir singular (buruk). Kondisi ini disebut sebagai *ill conditioned* (Draper dan Smith, 1992:299). Kondisi ini terjadi apabila terdapat korelasi antar variabel *regressor* yang cukup tinggi sehingga menyebabkan determinan $\mathbf{X}^T\mathbf{X}$ mendekati nol. Maka antara variabel *regressor* terjadi multikolinearitas ganda tidak sempurna.

Apabila terjadi situasi tersebut, penaksiran parameter koefisien regresi masih mungkin dilakukan, dengan metode kuadrat terkecil, tetapi dengan konsekuensi simpangan bakunya menjadi sangat sensitif sekalipun terjadi perubahan yang sangat kecil dalam datanya. Simpangan baku ini cenderung membesar sejalan dengan meningkatnya multikolinearitas.

Apabila terjadi multikolinearitas tidak sempurna pada variabel *regressor* pada diagonal utama $\mathbf{X}^T\mathbf{X}$ ditambah bilangan kecil positif β yang bernilai antara 0 dan 1 (Hoerl A.E, 1962). Prosedur ini disebut *Ridge Trace*. Kemudian prosedur tersebut dikembangkan oleh A.E Hoerl dan Robert W Kennard (1970) dan Normon R. Draper dan Harry Smith (1992) dengan mentransformasikan matriks $\mathbf{X}^T\mathbf{X}$ menjadi matriks korelasi.

Jadi regresi *Ridge* mengurangi dampak multikolinearitas dengan menentukan penduga yang bias tetapi mempunyai varians yang lebih kecil dari varians penduga regresi linear ganda. Pemilihan tetapan bias merupakan masalah yang perlu diperhatikan. Tetapan bias yang diinginkan adalah

tetapan bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien yang relatif stabil.

Secara umum model *ridge regression* dapat dinyatakan sebagai berikut :

$$\tilde{y} = (\mathbf{XW}\tilde{\beta} + \varepsilon)[\mathbf{I} - (\tilde{\beta}^*)^{-1}\mathbf{X}^T\mathbf{W}]^{-1} \quad (2.15)$$

dan

$$\begin{aligned} E(\tilde{y}) &= E(\mathbf{XW}\tilde{\beta}\mathbf{I}^{-1} - \mathbf{XW}\tilde{\beta}(\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} + \varepsilon\mathbf{I}^{-1} - \varepsilon((\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} \\ &= E(\mathbf{XW}\tilde{\beta}\mathbf{I}^{-1} + \varepsilon\mathbf{I}^{-1} - \mathbf{XW}\tilde{\beta}(\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} - \varepsilon((\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} \\ &= E(\mathbf{XW}\tilde{\beta}\mathbf{I}^{-1} + \mathbf{I}^{-1}E(\varepsilon)) - E(\mathbf{XW}\tilde{\beta}(\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} \\ &\quad - ((\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1}E(\varepsilon) \\ &= (\mathbf{XW}\tilde{\beta}\mathbf{I}^{-1}) - E(\mathbf{XW}\tilde{\beta}(\beta^*)^{-1}\mathbf{X}^T\mathbf{W})^{-1} \end{aligned}$$

2.5. Kajian Masalah Regresi dan *Outlier* dalam Al-Qur'an

2.5.1. 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 mempartisiny (membagi) dan hasil partisian ayat tersebut dimisalkan dengan sebuah variabel, yaitu :

$$\begin{aligned} \text{ذَٰلِكَ الْكِتَابُ لَا رَيْبَ ۚ فِيهِ هُدًى لِّلْمُتَّقِينَ ﴿٢﴾} \\ \text{الَّذِينَ يُؤْمِنُونَ} \\ \text{بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنْفِقُونَ ﴿٣﴾} \end{aligned}$$

Artinya : 2. Kitab (Al Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertaqwa,

3. (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezki yang Kami anugerahkan kepada mereka.

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

(Y)..... هُدًى لِّلْمُتَّقِينَ

(X)..... وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ

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

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.

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-Qomar 52 :

وَكُلُّ شَيْءٍ فَعَلُوهُ فِي الزُّبُرِ

Artinya : *"Dan segala sesuatu yang telah mereka perbuat tercatat dalam buku-buku catatan"*.

2.5.2. Kajian Keislaman tentang Outlier

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

Salah satu masalah ilmu statistika yang dibahas dalam penelitian ini adalah tentang *outlier* dalam ilmu statistik yang ternyata telah disinggung sejak zaman Nabi Muhammad. Hal

tersebut terbukti sebagaimana yang telah dijelaskan dalam surat Al-Fatihah ayat 7. Surat Al-Fatihah terdiri atas 7 ayat. Surat Al-Fatihah (Pembukaan) yang diturunkan di Mekah adalah surat yang pertama-tama diturunkan dengan lengkap di antara surat-surat yang ada dalam Al Quran dan termasuk golongan surat Makkiyyah. Surat ini disebut Al-Fatihah (Pembukaan), karena dengan surat inilah dibuka dan dimulainya Al-Quran. Dinamakan *Ummul Quran* (induk Al-Quran) atau *Ummul Kitaab* (induk Al-Kitab) karena dia merupakan induk dari semua isi Al Quran, dan karena itu diwajibkan membacanya pada tiap-tiap sembahyang. Dinamakan pula *As Sab'ul matsaany* (tujuh yang berulang-ulang) karena ayatnya tujuh dan dibaca berulang-ulang dalam sembahyang.

صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ

Artinya : “(yaitu) jalan orang-orang yang telah Engkau beri nikmat kepada mereka; bukan (jalan) mereka yang dimurkai dan bukan (pula jalan) mereka yang sesat.”

Yang dimaksud dengan mereka yang dimurkai dan mereka yang sesat ialah semua golongan yang menyimpang dari ajaran Islam. Asal turunnya Surat Al-Fatihah Ayat 7 yaitu untuk menampik dugaan bahwa semua umat Islam baik yang mendengar langsung ayat-ayat Al-Quran maupun yang belum atau tidak mendengarnya kesemuanya telah patuh kepada Allah. Kemudian pada ayat tersebut diterangkan bahwa (yaitu) jalan orang-orang yang telah Engkau beri nikmat kepada mereka yakni yang benar-benar taat dan penuh kepatuhan kepada Allah yang akan diberi nikmat dan ada pula

mereka yang sesat yakni mereka yang telah sangat jauh dari kebenaran lagi sangat mantap kekufurannya.

Kata “penyimpangan” dalam surat di atas pada konsep statistika dapat diartikan sebagai suatu *outlier*. Sebab suatu *outlier* dikatakan sebagai penyimpang dilihat dari pengertiannya 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 lagi dari rata-rata sisaannya.
2. *Outlier* adalah suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya (Draper dan Smith, 1992:146).
3. *Outlier* adalah data yang tidak mengikuti pola umum model (Sembiring, 1995:62).

Dari penafsiran Surat Al-Fatihah ayat 7 di atas dijelaskan bahwa “para penyimpang” yakni mereka yang telah sesat, yang sangat jauh dari kebenaran lagi sangat mantap kekufurannya. Penafsiran mengenai para penyimpang tersebut mempunyai makna yang sama dengan pengertian dari *outlier* yaitu sama-sama terletak sangat jauh.

BAB III

PEMBAHASAN

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

3.1. Menentukan Model Regresi Linier Terboboti dengan Metode Ridge

Diasumsikan model regresi linier terboboti yang dipakai adalah model regresi linier terboboti yang mengandung *outlier* sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} W_{1i} + \beta_2 X_{2i} W_{2i} + \cdots + \beta_k X_{kn} W_{kn} + \varepsilon_i \quad (3.1)$$

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

Dalam bentuk matriks persamaan di atas menjadi :

$$\begin{aligned} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} &= \begin{bmatrix} 1 & X_{11}W_{11} & X_{21}W_{21} & \cdots & X_{k1}W_{k1} \\ 1 & X_{12}W_{12} & X_{22}W_{22} & \cdots & X_{k2}W_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n}W_{1n} & X_{2n}W_{2n} & \cdots & X_{kn}W_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k1} & W_{k2} & \cdots & W_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (3.2) \end{aligned}$$

Jika dimisalkan :

$$\begin{aligned} Y^* &= \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} & X &= \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \\ \beta^* &= \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} & W^* &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k1} & W_{k2} & \cdots & W_{kn} \end{bmatrix} \end{aligned}$$

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

Persamaan (3.2) dapat dinyatakan sebagai berikut :

$$Y^* = X^*W\tilde{\beta}^* + \varepsilon^* \quad (3.3)$$

dimana :

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

X^* : matriks peubah bebas yang mengandung *outlier* ukuran $n \times (k+1)$

β^* : vektor parameter model regresi *Ridge* ukuran $(k+1) \times 1$ yang tak diketahui

W : matriks pembobot ukuran $(k+1) \times (k+1)$

ε^* : vektor *error* ukuran $n \times 1$

dimana matriks W adalah matriks pembobot ukuran $(k+1) \times (k+1)$ yang digunakan untuk memboboti X pada model regresi terboboti.

maka persamaan (3.3) dapat dinyatakan dengan :

$$Y_{nx1}^* = X_{nx(k+1)}^* W_{(k+1)(k+1)} \tilde{\beta}_{(k+1)x1}^* + \varepsilon_{nx1}^* \quad (3.4)$$

Setelah didapatkan persamaan pada model (3.4) maka hasil selanjutnya untuk mengatasi apakah *outlier* pada $\tilde{\beta}^*$ sudah hilang adalah dengan cara mencari estimasi parameter dari $\tilde{\beta}^*$.

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

3.2.1. Estimasi Parameter $\tilde{\beta}^*$

Dari persamaan (3.4) diketahui bahwa Untuk mengetahui apakah model (3.4) mengandung *outlier*, maka parameter $\tilde{\beta}$ dari

model Ridge dicari nilai kuadrat *error*nya dengan cara sebagai berikut :

$$\tilde{\varepsilon} = \tilde{y}(\mathbf{I} - \tilde{\beta}^*(\mathbf{X}^T\mathbf{W})) - \mathbf{XW}\tilde{\beta}$$

$$\tilde{\varepsilon} = \tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta}$$

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = (\tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta})^T (\tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta}) \quad (3.5)$$

dengan

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = S \quad (3.6)$$

atau S dapat dijabarkan sebagai berikut :

$$\begin{aligned} S &= (\tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta})^T (\tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta}) \\ &= (\tilde{y}^T - (\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T - \tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T) (\tilde{y} - \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \mathbf{XW}\tilde{\beta}) \\ &= \tilde{y}^T \tilde{y} - 2\tilde{y}^T \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - \tilde{y}^T \mathbf{XW}\tilde{\beta} + (\mathbf{WX}^T)^T \tilde{\beta}^{*T} \tilde{y}^T \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) \\ &\quad + (\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW}\tilde{\beta} - (\tilde{y}^T \mathbf{XW}\tilde{\beta}) + (\mathbf{X}^T\mathbf{W})^T \tilde{\beta}^{*T} \tilde{y}^T \mathbf{XW}\tilde{\beta} \\ &\quad + \tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta} \\ &= \tilde{y}^T \tilde{y} - 2\tilde{y}^T \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - 2\tilde{y}^T \mathbf{XW}\tilde{\beta} + (\mathbf{WX}^T)^T \tilde{\beta}^{*T} \tilde{y}^T \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) \\ &\quad + (\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW}\tilde{\beta} + (\mathbf{X}^T\mathbf{W})^T \tilde{\beta}^{*T} \tilde{y}^T \mathbf{XW}\tilde{\beta} + \tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta} \\ &= \tilde{y}^T \tilde{y} - 2\tilde{y}^T \tilde{y}\tilde{\beta}^*(\mathbf{X}^T\mathbf{W}) - 2\tilde{y}^T \mathbf{XW}\tilde{\beta} + (2\mathbf{WX}^T)^T \tilde{\beta}^{*T} \tilde{y}^T \tilde{y}\tilde{\beta}^* \\ &\quad (\mathbf{X}^T\mathbf{W}) + \tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T \mathbf{XW}\tilde{\beta} \end{aligned} \quad (3.7)$$

Untuk mengestimasi parameter $\tilde{\beta}$ yang dinotasikan dengan $\tilde{\beta}^*$, adalah dengan mendeferensialkan persamaan (3.7) terhadap $\tilde{\beta}^*$ dan disamadengankan nol.

Dideferensialkan terhadap $\tilde{\beta}^*$:

$$\begin{aligned}
\frac{dS}{d\tilde{\beta}^*} &= 0 - (2\tilde{y}^T \tilde{y}(\mathbf{X}^T \mathbf{W}))^T - (2\tilde{y}^T \mathbf{XW})^T + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} \\
&\quad + ((2\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW})^T + \mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta} + (\tilde{\beta}^T \mathbf{W}^T \mathbf{X}^T \mathbf{XW})^T \\
&= -2(\mathbf{X}^T \mathbf{W})^T \tilde{y}^T \tilde{y} - 2\mathbf{W}^T \mathbf{X}^T \tilde{y} + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} + \\
&\quad \mathbf{W}^T \mathbf{X}^T \tilde{y} \tilde{\beta} (2\mathbf{WX}^T) + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta} \\
&= -2(\mathbf{X}^T \mathbf{W})^T \tilde{y}^T \tilde{y} - 2\mathbf{W}^T \mathbf{X}^T \tilde{y} + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} + \\
&\quad (2\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW} + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta} \tag{3.7}
\end{aligned}$$

Kemudian turunan keduanya adalah sebagai berikut,

$$\begin{aligned}
\frac{dS}{d\tilde{\beta}^*} &= -2(\mathbf{X}^T \mathbf{W})^T \tilde{y}^T \tilde{y} - 2\mathbf{W}^T \mathbf{X}^T \tilde{y} + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} \\
&\quad + (2\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW} + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta} \\
&= 0 - 0 + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} + ((2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW})^T + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \\
&= (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} + ((2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW})^T + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW}
\end{aligned}$$

Kemudian persamaan di atas disamadengankan nol

$$\begin{aligned}
0 &= -2(\mathbf{X}^T \mathbf{W})^T \tilde{y}^T \tilde{y} - 2\mathbf{W}^T \mathbf{X}^T \tilde{y} + (2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} \\
&\quad + (2\mathbf{WX}^T)^T \tilde{\beta}^T \tilde{y}^T \mathbf{XW} + 2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta} \\
&\quad \left((2\mathbf{WX}^T)^T \tilde{y}^T \mathbf{XW} \tilde{\beta} + (2\mathbf{W}^T \mathbf{X}^T \mathbf{XW} \tilde{\beta}) \right) = \\
&\quad -2(\mathbf{X}^T \mathbf{W})^T \tilde{y}^T \tilde{y} - 2\mathbf{W}^T \mathbf{X}^T \tilde{y} + (\mathbf{W}^T \mathbf{X}^T \tilde{y} (2\mathbf{WX}^T))^{-1} \tag{3.8}
\end{aligned}$$

Dari persamaan (3.8) maka didapatkan $\tilde{\beta}^*$ sebagai berikut :

$$\tilde{\beta}^* = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \tilde{\beta}_{WLS} \mathbf{I}_r)^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{y} \tag{3.9}$$

Jadi estimasi dari parameter $\hat{\beta}^*$ adalah persamaan (3.9) yaitu :

$$\tilde{\beta}^* = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \tilde{\beta}_{WLS} \mathbf{I}_r)^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{y} \tag{3.10}$$

Estimasi parameter pada persamaan (3.10) dikatakan sebagai estimasi parameter $\hat{\beta}^*$ yang mengandung *outlier*. Setelah didapatkan $\tilde{\beta}^*$, maka selanjutnya dicari sifat-sifat estimasi dari parameter $\tilde{\beta}^*$ tersebut.

3.3. Menentukan Sifat-Sifat Estimasi Parameter Regresi Linier yang Mengandung *Outlier*

Salah satu cara menentukan sifat-sifat estimasi model regresi linier terboboti yang mengandung *outlier* adalah dengan menentukan sifat-sifat dari parameter $\tilde{\beta}^*$.

3.3.1. Tak Bias (*Unbias*)

$\tilde{\beta}$ dikatakan estimator tak bias jika $E(\hat{\beta}^*) = \tilde{\beta}^*$

Bukti :

$$\begin{aligned}
 E(\hat{\beta}^*) &= E(\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} \tilde{\mathbf{y}} \\
 &= E(\mathbf{X}^T \mathbf{W} \mathbf{X} + ((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}) \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) E(\tilde{\mathbf{y}})) \\
 &= E(\mathbf{X}^T \mathbf{W} \mathbf{X} + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}) \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) (\mathbf{X} \mathbf{W} \tilde{\beta}) \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}) \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) (\mathbf{X} \mathbf{W} \tilde{\beta}) \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} + (\mathbf{X}^T \mathbf{W} \mathbf{X}) (\mathbf{X}^T \mathbf{W}) \mathbf{I})^{-1} \mathbf{X} \mathbf{W} \tilde{\beta} \\
 &= ((\mathbf{X}^T \mathbf{W}) \mathbf{I})^{-1} \mathbf{X} \mathbf{W} \tilde{\beta} \\
 &= \mathbf{I} \tilde{\beta}^* \\
 &= \tilde{\beta}^*
 \end{aligned} \tag{3.10}$$

Dari persamaan (3.10) diperoleh $E(\hat{\beta}^*) = \tilde{\beta}^*$ maka $\tilde{\beta}^*$ yang mengandung *outlier* merupakan estimator tak bias.

3.3.2. Efisien

Suatu estimator dikatakan efisien apabila estimator tersebut mempunyai variansi kecil.

Perhatikan bahwa :

$$\begin{aligned} Var(\hat{\beta}^*) &= E \left[\left(\hat{\beta}^* - E(\hat{\beta}^*) \right) \left(\hat{\beta}^* - E(\hat{\beta}^*) \right)^T \right] \\ &= \left[(\hat{\beta}^* - \tilde{\beta}^*) (\hat{\beta}^* - \tilde{\beta}^*)^T \right] \end{aligned} \quad (3.11)$$

Karena

$$\begin{aligned} \hat{\beta}^* &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} \tilde{\mathbf{y}} \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} (\mathbf{X} \mathbf{W} \tilde{\beta} + \tilde{\varepsilon}^*) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} \left((\mathbf{X} \mathbf{W}) \tilde{\beta} + \tilde{\varepsilon}^* \right) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \left((\mathbf{X}^T \mathbf{W}) (\mathbf{X} \mathbf{W}) \tilde{\beta} + (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \right) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \left((\mathbf{X}^T \mathbf{X}) (\mathbf{X} \mathbf{W}) (\mathbf{X}^T \mathbf{W}) \tilde{\beta} + (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \right) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \left((\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{W}) \tilde{\beta} + (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \right) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{W}) \tilde{\beta} \\ &\quad + (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \\ &= \mathbf{I} \beta^* + (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \\ &= \beta^* + (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \\ \hat{\beta}^* - \beta^* &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \end{aligned}$$

maka

$$\begin{aligned}
 \text{Var}(\hat{\beta}^*) &= E \left[(\hat{\beta}^* - E(\hat{\beta}^*)) (\hat{\beta}^* - E(\hat{\beta}^*))^T \right] \\
 &= E [((\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^*) ((\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^*)^T] \\
 &= E [(\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{\varepsilon}^* \tilde{\varepsilon}^{*T} (\mathbf{X}^T \mathbf{W}) (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1}] \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) E [\tilde{\varepsilon}^* \tilde{\varepsilon}^{*T}] (\mathbf{X}^T \mathbf{W}) (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) \sigma^2 \mathbf{I} (\mathbf{X}^T \mathbf{W}) (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}) (\mathbf{X} \mathbf{W}) (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \sigma^2 \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{W}) (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \sigma^2 \\
 &= \mathbf{I} (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \sigma^2 \\
 &= (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \sigma^2 \tag{3.12}
 \end{aligned}$$

Sehingga $\text{Var}(\hat{\beta}^*) = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \beta^* \mathbf{I})^{-1} \sigma^2$ harus sekecil mungkin agar $\hat{\beta}^*$ efisien.

3.3.3. Konsisten

Estimator yang konsisten adalah

$$E \left(\hat{\beta} - E(\beta) \right)^2 \rightarrow 0 \text{ jika } n \rightarrow \infty$$

Sehingga

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

Dari persamaan (3.10) diperoleh $E(\hat{\beta}^*) = \tilde{\beta}^*$ maka

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

$$\begin{aligned}
&= (\beta^* - \beta^*)(\hat{\beta}^* - \beta^*)^T \\
&= (0)(\hat{\beta}^* - \beta^*)^T \\
&= 0
\end{aligned} \tag{3.13}$$

Dari persamaan (3.13) diperoleh $E \left[(\hat{\beta}^* - E(\hat{\beta}^*)) (\hat{\beta}^* - E(\hat{\beta}^*))^T \right] = 0$, maka untuk $\hat{\beta}^*$ yang mengandung *outlier* merupakan estimator yang konsisten.

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

3.4.1. Diskripsi Data

Data berikut berasal dari Atkinson A.C (1981: 281), data ini adalah data 25 pasien dari sebuah rumah sakit. Data ini merupakan data tes fungsi enzyme. Datanya adalah sebagai berikut :

Tabel 3.1 Data Tes Fungsi Enzim

No.	Y	X1	X2	X3	X4
1	200	6.7	62	81	2.59
2	101	5.1	59	66	1.7
3	204	7.4	57	83	2.16
4	101	6.5	73	41	2.01
5	509	7.8	65	115	4.3
6	80	5.8	38	72	1.42
7	80	5.7	46	63	1.91
8	127	3.7	68	81	2.57
9	202	6	67	93	2.5
10	203	3.7	76	94	2.4
11	329	6.3	84	83	4.13
12	65	6.7	51	43	1.86
13	330	5.8	83	88	3.95
14	168	7.7	62	67	3.4
15	217	7.4	74	68	2.4
16	87	6	85	28	2.98
17	34	3.7	51	41	1.55
18	215	7.3	68	74	3.56
19	172	5.6	57	87	3.02
20	109	5.2	52	76	2.85
21	830	5.8	96	114	3.95
22	136	3.4	83	53	1.12
23	70	6.7	26	68	2.1
24	220	5.8	67	86	3.4
25	276	6.3	59	100	2.95

Sumber : Atkinson A.C (1981). *Two Graphical Display for Outlying and Influential Observation in Regression. Technometrics*

Keterangan : X_1 = Blood Clotting Score;

X_2 = Prognostic Index;

X_3 = Enzyme Function Test;

X_4 = Liver Function Test;

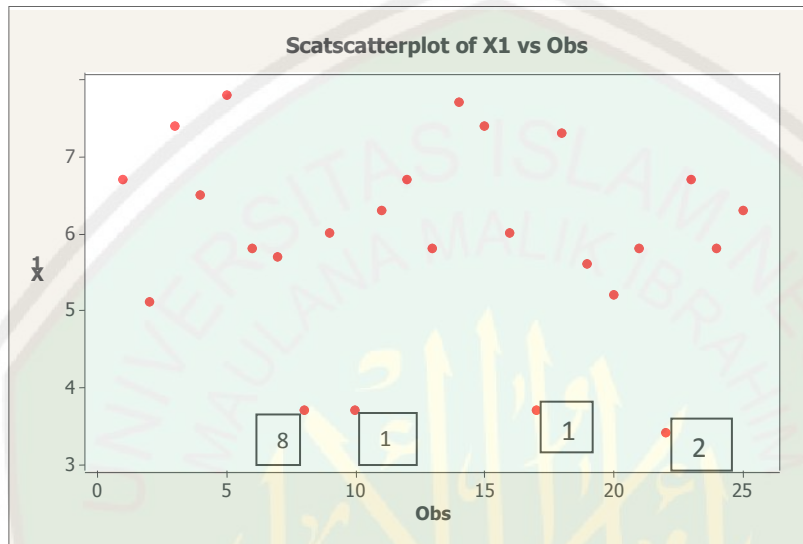
Y = Survival Time

1.4.2 Metode Grafis

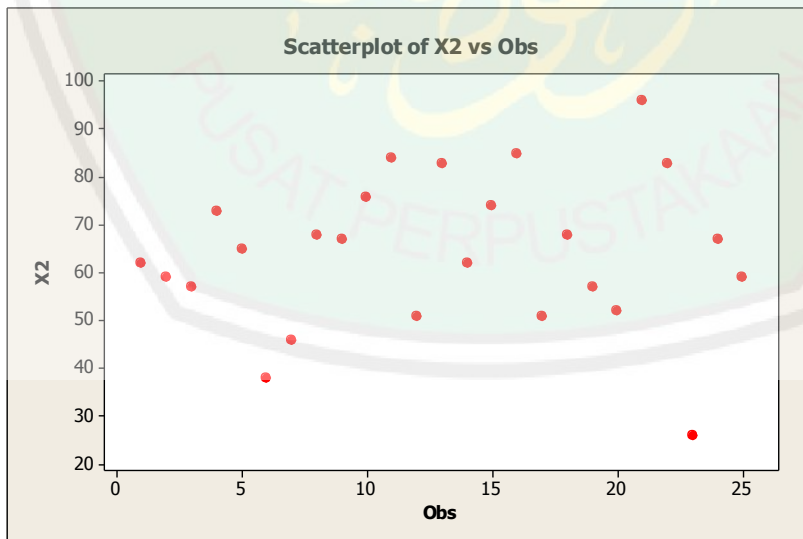
Berdasarkan data di atas, akan dilakukan pengidentifikasian keberadaan

outlier, yaitu :

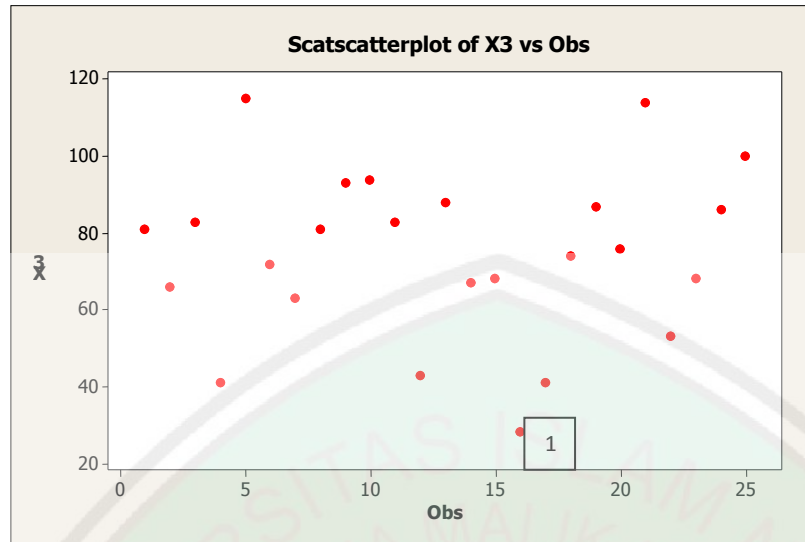
Melalui metode grafis diperoleh gambar berikut :



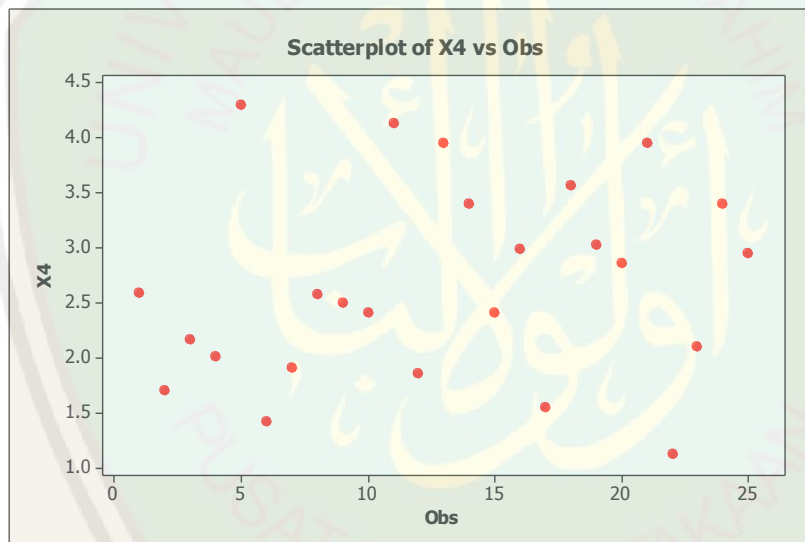
Gambar 3.1 Scatterplot of X1



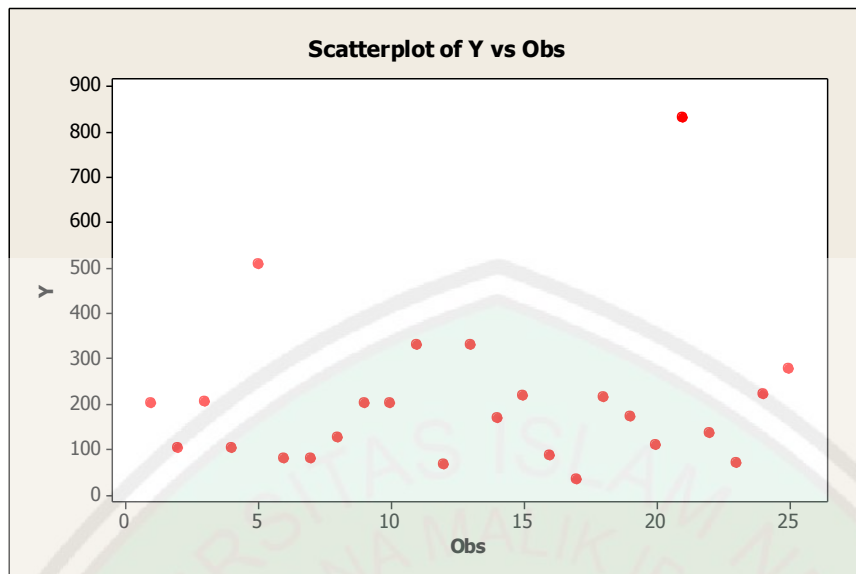
Gambar 3.2 Scatterplot of X2



Gambar 3.3 Scatterplot of X3



Gambar 3.4 Scatterplot of X4



Gambar 3.5 Scatterplot of Y

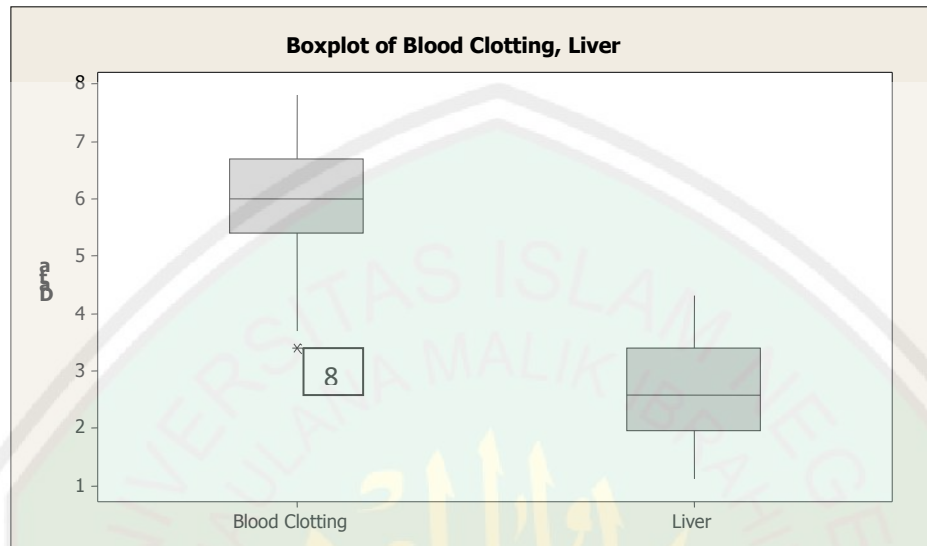
Kelima gambar di atas menunjukkan *scatter-plot* untuk semua variabel penelitian. Pada masing-masing plot mengindikasikan adanya outlier. Demi memudahkan pembacaan gambar, diberikan label nomor pada data yang diduga merupakan outlier.

Untuk lebih memberikan keyakinan atas keputusan dari hasil analisis visual di atas, dilakukan analisis lainnya.

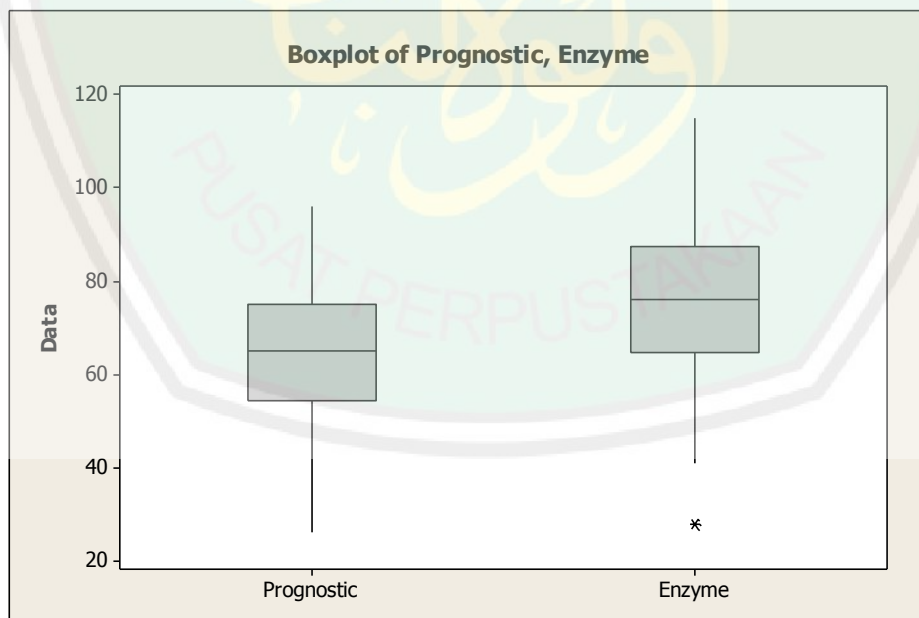
1.4.3 Boxplot

Untuk lebih memperjelas pemahaman mengenai outlier ini, teori-teori yang telah diuraikan sebelumnya akan diterapkan pada gambar Boxplot berikut. Berdasarkan data di atas, akan dilakukan pengidentifikasian keberadaan outlier. Gambar di bawah ini menunjukkan boxplot untuk semua variabel penelitian. Pada masing-masing plot mengindikasikan adanya outlier. Demi memudahkan pembacaan gambar, penulis memberikan label nomor pada data yang

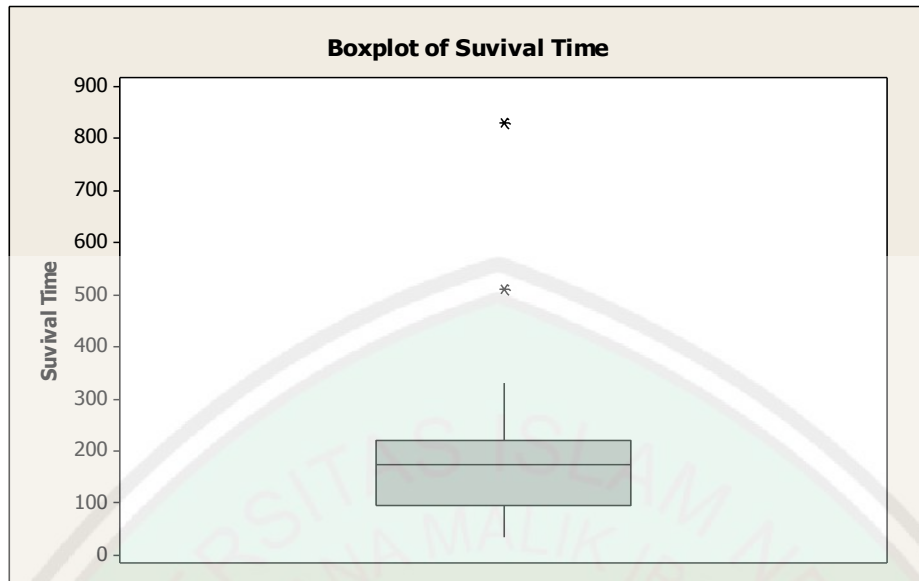
diduga merupakan outlier. Bila disajikan dalam boxplot akan tampak seperti di bawah ini :



Gambar 3.6 Boxplot of X1, X2



Gambar 3.7 Boxplot of X3, X4



Gambar 3.8 Boxplot of Y

Berdasarkan ketiga boxplot di atas, diketahui terdapat beberapa pencilan pada masing-masing variabel, yaitu :

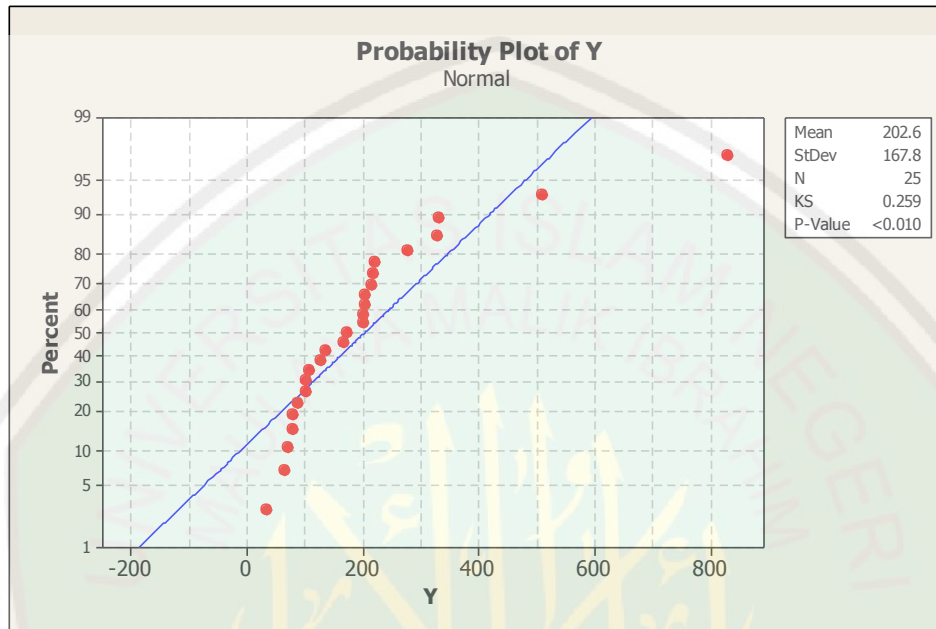
- Data ke-8, 10, 17 dan 22 pada variabel *Blood Clotting* (X_1)
- Data ke-16 pada variabel *Enzyme* (X_3)

3.4.4 Penanggulangan *Outlier*

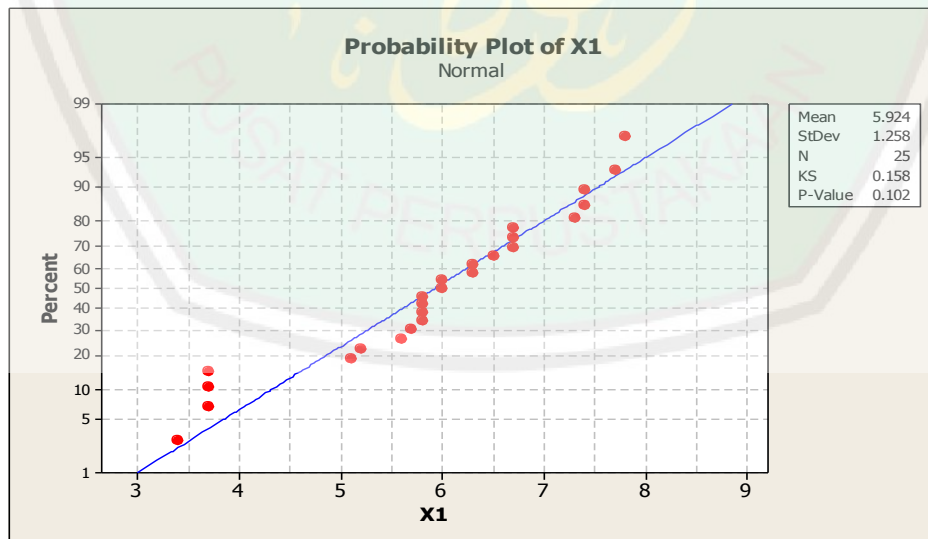
Untuk menanggulangi *outlier* pada data, yaitu dengan mengeluarkan atau membuang observasi ke-I pada data yang diduga merupakan *outlier*. Kemudian dilakukan pengujian kembali untuk mendeteksi terdapat atau tidaknya *outlier* pada data sampai tidak terdapat lagi *outlier* pada data tersebut.

Meskipun *outlier* identik dengan data yang tidak bagus, akan tetapi ia merupakan bagian terpenting dari data, karena menyimpan informasi tertentu.

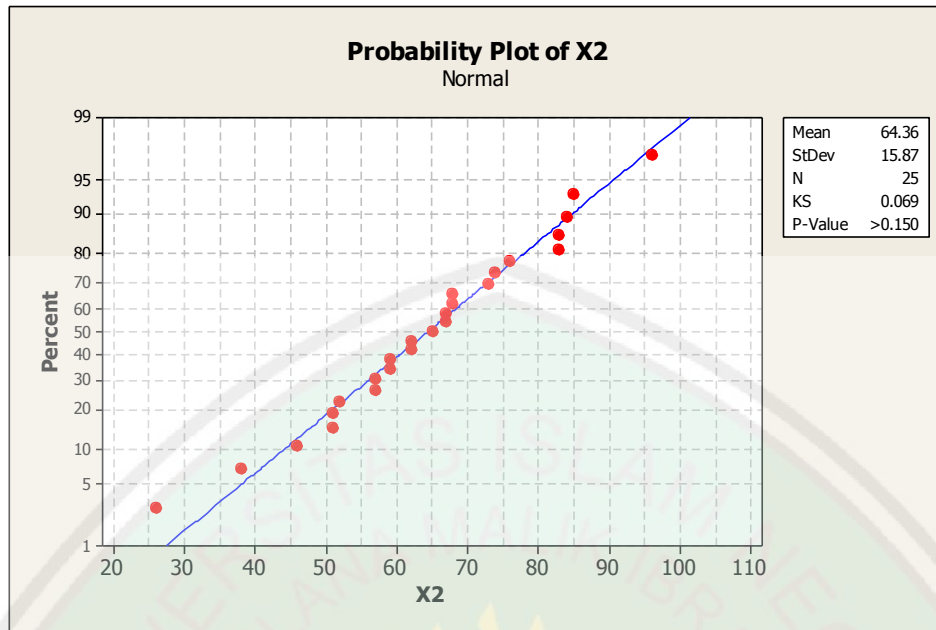
Untuk itu, alternatif yang dapat diambil terhadap data yang terkontaminasi outlier adalah dengan menggunakan metode *Ridge* dalam penaksiran model regresi.



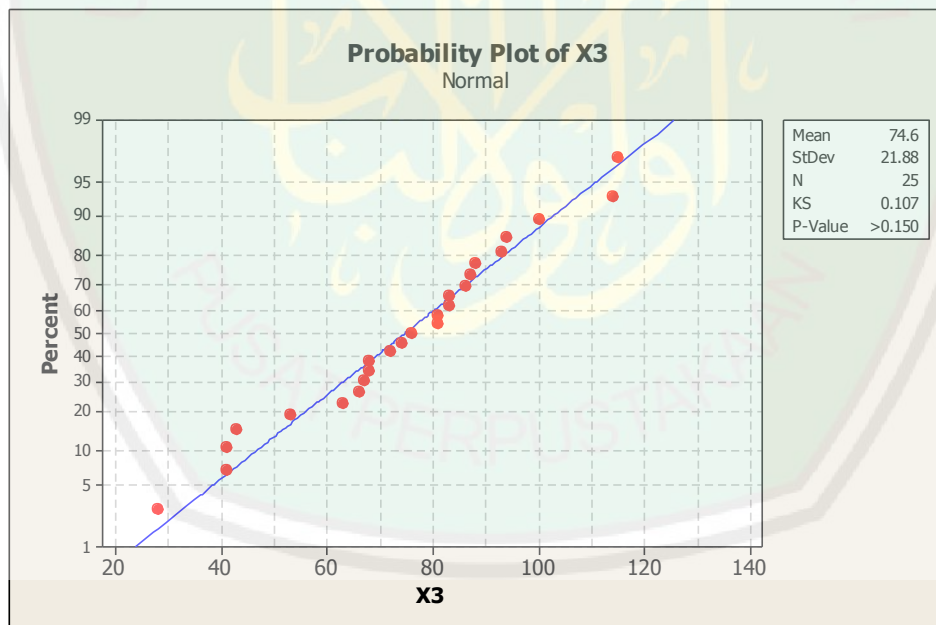
Gambar 3.9 Probability Plot of Y



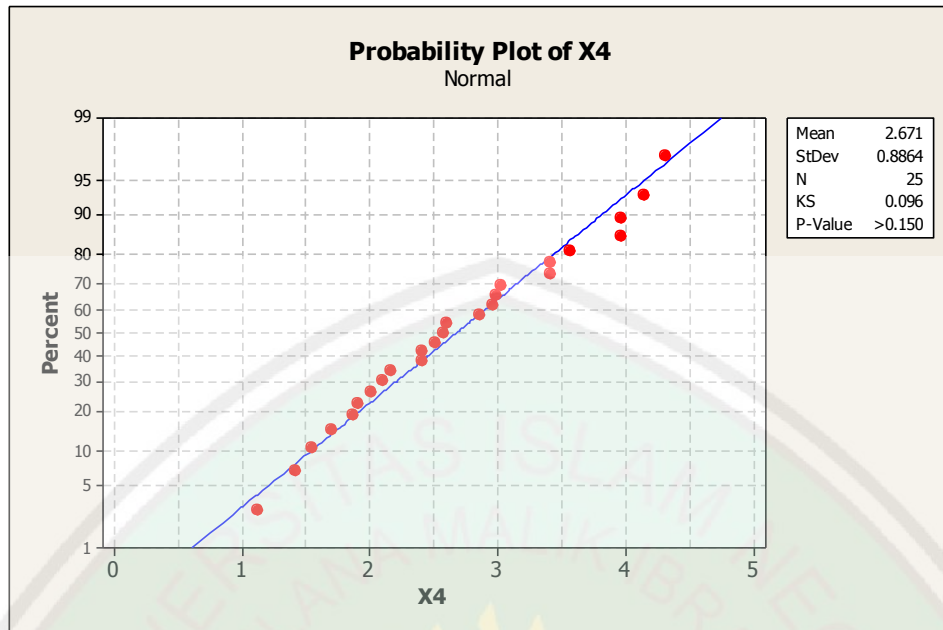
Gambar 3.10 Probability Plot of X1



Gambar 3.11 Probability Plot of X2



Gambar 3.12 Probability Plot of X



Gambar 3.13 Probability Plot of X4

Regression Analysis: Y versus X1, X2, X3, X4

The regression equation is

$$Y = -625 + 23.9 X1 + 4.97 X2 + 4.60 X3 + 8.7 X4$$

Predictor	Coef	SE Coef	T	P
Constant	-625.1	138.6	-4.51	0.000
X1	23.86	17.13	1.39	0.179
X2	4.967	1.403	3.54	0.002
X3	4.604	1.027	4.48	0.000
X4	8.67	33.41	0.26	0.798

S = 85.3961 R-Sq = 78.4% R-Sq(adj) = 74.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	529988	132497	18.17	0.000
Residual Error	20	145850	7292		
Total	24	675838			

Source DF Seq SS

X1	1	35941
X2	1	248482
X3	1	245073
X4	1	492

Unusual Observations

Obs	X1	Y	Fit	SE Fit	Residual	St Resid
21	5.80	830.0	549.3	47.0	280.7	3.94R

R denotes an observation with a large standardized residual.

Gambar 3.14 Hasil Analisis Regresi Linier dengan Minitab

3.5 Keterkaitan Hasil Penelitian dengan Kajian Agama

Pada BAB II telah disinggung bahwa regresi terdapat pada surat Al-Baqarah ayat 2-3. Peneliti pada bab ini akan menghubungkan antara Qs. 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 الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ ﴿٢﴾ الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ
وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ ﴿٣﴾

Artinya : 2. Kitab (Al Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertaqwa,
3. (yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezki 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 sebahagian rezki 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-fatihah ayat 7 sebagai berikut:

صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ ﴿٧﴾

Artinya : “(yaitu) jalan orang-orang yang telah Engkau beri nikmat kepada mereka; bukan (jalan) mereka yang dimurkai dan bukan (pula jalan) mereka yang sesat.”

Surat Al-fatihah ayat 7 di atas, Allah menjelaskan tentang sekumpulan orang-orang bahwa di antara mereka ada yang beriman dan diberi nikmat, khusyuk dan ikhlas serta beramal saleh karena-Nya. Ada pula di antara mereka yang berpaling dari haluan yang benar, dimurkai dan tersesat. Barangsiapa beriman kepada Allah dan mentaati-Nya sesungguhnya dia telah menempuh jalan yang akan menyampaikannya

kepada kebahagiaan dan telah melakukan sesuatu yang akan menyelamatkannya dari siksa neraka.

Setelah diuraikan di atas dapat diambil kesimpulan bahwa yang menjelaskan *outlier* adalah kalimat “(yaitu) *jalan orang-orang yang telah Engkau beri nikmat kepada mereka; bukan (jalan) mereka yang dimurkai dan bukan (pula jalan) mereka yang sesat*” dalam artian *outlier* adalah suatu yang menyimpang atau tersesat.

Dari penafsiran ayat ini dijelaskan bahwa para penyimpang 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.

BAB IV

PENUTUP

4.1. Kesimpulan

Dari metode *Ridge* pada pembahasan di bab III didapatkan estimasi parameter $\tilde{\beta}^*$ yaitu $\tilde{\beta}^* = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \tilde{\beta}_{WLS} \mathbf{I}_r)^{-1} (\mathbf{X}^T \mathbf{W}) \tilde{y}$ memenuhi syarat *unbias*. Sehingga $\tilde{\beta}^*$ yang didapat sudah dapat digunakan untuk mengatasi *outlier* pada model regresi terboboti.

4.2. Saran

Diharapkan untuk penelitian selanjutnya menggunakan estimator yang lain untuk mengatasi *outlier* untuk mencari estimasi parameternya dan juga dapat menggunakan metode yang lain selain metode *Ridge* untuk model regresi terboboti.

DAFTAR PUSTAKA

- Algifari. 1997. *Analisis Regresi, Teori Kasus dan Solusi*. Yogyakarta : BPFE.
- Atkinson A.C. 1981. Two Graphical Display for Outlying and Influential Observation in Regression. *Technometrics*.
- Cahyawati, Dian, dkk. Tanpa Tahun. *Efektivitas Metode Regresi Robust Penduga Welsch dalam Mengatasi Pencilan pada Pemodelan Regresi Linear Berganda*. Universitas Sriwijaya, Sumatera Selatan, Indonesia. Volume 12 nomer 1(A).
- Draper, N.R, & Smith, H. 1992. *Analisis Regresi Terapan*. Terjemah oleh Bambang Sumantri. Jakarta : PT Gramedia Pustaka Utama.
- Hasan, Iqbal. 2002. *Pokok-Pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta: Bumi Aksara.
- Hoerl, A.E dan Kennard, R.W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, Vol.12, no.1
- Sembiring, RK. 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](http://resources.unpad.ac.id/unpad-content/uploads/publikasi_dosen/Outlier(Pencilan).pdf) (diunduh pada tanggal 10 Juli 2012).