TAKSIRAN PARAMETER PADA MODEL REGRESI ROBUST DENGAN MENGGUNAKAN FUNGSI HUBER

STEVANI WIJAYA

030501061Y

UNIVERSITAS INDONESIA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

DEPARTEMEN MATEMATIKA

DEPOK

2009

TAKSIRAN PARAMETER PADA MODEL REGRESI ROBUST

DENGAN MENGGUNAKAN FUNGSI HUBER

Skripsi diajukan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Oleh:

STEVANI WIJAYA

030501061Y

DEPOK

2009

SKRIPSI : TAKSIRAN PARAMETER PADA MODEL REGRESI ROBUST DENGAN MENGGUNAKAN FUNGSI HUBER

NAMA : STEVANI WIJAYA

NPM : 030501061Y

SKRIPSI INI TELAH DIPERIKSA DAN DISETUJUI

DEPOK, JULI 2009

Dra. RIANTI SETIADI , M.Si.	Dra. SASKYA MARY , M.Si.
PEMBIMBING I	PEMBIMBING II

Tanggal lulus Ujian Sidang Sarjana:

Penguji I :

Penguji II :

Penguji III :

KATA PENGANTAR

Puji syukur kepada Tuhan atas segala rahmat, berkat dan kekuatan yang telah diberikan sehingga penulis dapat menyelesaikan tugas akhir ini. Penyelesaian tugas akhir ini tidak terlepas dari dukungan dan doa dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada berbagai pihak sebagai berikut :

1. Keluarga tercinta, papa, mama, ce Kaka, Ade dan ii Kiang2 yang telah memberikan segala bentuk dukungan, doa dan motivasi untuk terus tetap bertahan dalam menghadapi segala masalah yang dihadapi saat menyelesaikan tugas akhir ini. Maaf atas segala keluh kesah penulis.
2. Pembimbing tugas akhir penulis, Dra. Rianti Setiadi, M.Si dan Dra. Saskya Mary, M.Si yang telah menyediakan banyak waktu, tenaga, pikiran serta dukungan mental kepada penulis. U're the best!!!
3. Pembimbing akademik penulis, Ibu Rustina, para dosen serta staf Departemen Matematika UI yang telah membantu selama kuliah .
4. Sahabat-sahabat penulis, Jessie O, Maria W, Clara Vania, Priskilla Pratita, Alberta (Mei), Martinus P, Hadi GS, Andrea P, Daniel H, Gayatri, Cicilia, Sherly N, Maria Theodora, karena kalian penulis bertahan sampai akhir. Terima kasih telah mendengar semua keluh kesah penulis, memberi kekuatan dan keyakinan. *(Gaya : Apapun yang terjadi, itu rencana Dia!)*

5. Teman-teman seperjuangan yang mengambil skripsi, Wakhidah , Mayramadan, Amri, Khuriyanti, Shinta , Ratih, Rizky, Rifky, Riesa, Syarah, Maul, Uun, k Iif. Temanz, perjuangan kita tidak sia-sia.. :p
6. Romo Markus Yumartana, Damianus Fritz, Irwanto atas kekuatan, nasehat, motivasi dan keyakinan yang diberikan. *(Rm Yu : Mazmur 126:5, kak Fritz : kamu tidak sendiri dek, banyak yang mendoakanmu!)*
7. Theja Salim dan Ponco Ridwan atas saran dan bantuan yang telah diberikan. *(k Theja : bukunya benar2 membantu..thx a lot)*.
8. Semua Keluarga Mahasiswa Katolik FMIPA UI, Anggha, Inne, Nteph, Ranti, Pangky, dan yang lainnya yang tidak dapat disebutkan semuanya yang telah memberikan doa dan semangat.
9. Teman-teman angkatan 2005, Ratna, Melati, Raisa, Nisma, Othe, Miranti, Rani, Fika, Anggie, Akmal, Anggie, Wicha, Dia, Puji, Shally, Gyo, Pute, Aini, Rif'ah, Rara, Yanu, Merry, Yuni, Fia, Dian, Mia, Hamdan, Asep, Trian, Ridwan, Aris, Hairu. Thx atas smuanya.
10. Semua pihak yang telah membantu namun tidak disebutkan satu persatu karena keterbatasan tempat.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, akhir kata penulis mengucapkan banyak maaf atas semua kesalahan, semoga tugas akhir ini dapat bermanfaat bagi banyak orang.

Depok, Juli 2009

Penulis

ABSTRAK

Dalam analisis data, saat data mempunyai *outlier* dan *outlier* yang ada bukan merupakan suatu kesalahan, taksiran parameter yang diperoleh dengan metode *Ordinary Least Square* (OLS) akan bias karena metode OLS tidak *robust* terhadap adanya *outlier*. Oleh karena itu, dicari metode lain yang *robust* terhadap adanya *outlier*, salah satunya ialah metode regresi robust dengan menggunakan fungsi Huber. Pada skripsi ini akan dibahas mengenai taksiran parameter pada model regresi robust sederhana dan berganda dengan menggunakan fungsi Huber. Selain itu, akan dibandingkan antara taksiran parameter model regresi robust dengan menggunakan fungsi Huber dan taksiran parameter yang didapat dengan metode OLS dilihat dari nilai effisiensi taksiran parameter. Hasil yang diperoleh dari contoh penerapan menunjukkan bahwa untuk data ada *outlier* taksiran parameter yang diperoleh dengan metode regresi robust dengan fungsi Huber lebih effisien dibandingkan metode OLS, sedangkan untuk data tanpa *outlier* taksiran parameter yang diperoleh dengan metode OLS lebih effisien dibandingkan metode regresi robust dengan fungsi Huber.

Kata Kunci : effisien, fungsi Huber, metode OLS, *outlier*, regresi robust

x + 100 hlm.; lamp.; tab.

Bibliografi : 13 (1980-2008)

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk melihat hubungan fungsional antara variabel dependen (variabel terikat) dan variabel independen (variabel bebas) biasanya digunakan model regresi. Salah satu cara untuk menaksir parameter pada model regresi yang sering digunakan adalah dengan metode *Ordinary Least Square* (OLS) yang meminimumkan jumlah kuadrat *error*. Karena taksiran dengan metode OLS dicari berdasarkan jumlah kuadrat *error*, maka adanya *outlier* akan menyebabkan taksiran parameter yang didapat menjadi bias. Yang dimaksud dengan *outlier* di sini adalah suatu observasi yang menyimpang jauh dari hubungan linear yang dibentuk oleh mayoritas dari data. Dalam hal ini variabel independen dan variabel dependen diperhitungkan secara simultan.

Pada analisis regresi linier dengan metode OLS, *outlier* dicari berdasarkan suatu besaran yang dihitung dari data. Ada banyak cara untuk mendapatkan besaran tersebut seperti telah dibahas dalam banyak tulisan tentang model regresi. Setelah *outlier* ditemukan maka *outlier* tersebut dikeluarkan dari analisis dan kemudian taksiran parameter dicari dengan metode OLS berdasarkan data tanpa *outlier*. Hal ini tidak bermasalah jika *outlier* berasal dari suatu kesalahan. Akan tetapi, jika *outlier* bukan berasal dari kesalahan, tindakan mengeluarkan *outlier* dari analisis bukan merupakan tindakan yang tepat karena *outlier* dapat memberikan informasi yang justru tidak diberikan oleh observasi lainnya. Selain itu mengeluarkan *outlier* dari analisis dapat membuat taksiran parameter yang diperoleh menjadi “*under-estimate*”.

Salah satu metode untuk mengatasi masalah *outlier* adalah dengan menggunakan regresi robust. Pada regresi robust, taksiran yang *robust* terhadap *outlier* (tidak terpengaruh oleh adanya *outlier*) akan dicari sehingga *outlier* yang ada tidak perlu dikeluarkan dari analisis. Ada beberapa cara untuk mendapatkan taksiran yang *robust* terhadap *outlier*, salah satunya adalah taksiran *robust* dengan menggunakan fungsi Huber. Pada prinsipnya, metode penaksiran dengan menggunakan fungsi Huber akan mencari taksiran parameter dengan meminimumkan total fungsi dari *error*.

Pada tugas akhir ini, penulis akan membahas tentang taksiran parameter pada model regresi robust dengan menggunakan fungsi Huber. Bahasan akan mencakup tentang bagaimana mencari taksiran parameter pada model regresi robust dengan menggunakan fungsi Huber, baik pada model regresi robust sederhana maupun pada model regresi robust berganda. Dengan contoh data, taksiran parameter yang didapat dengan metode regresi robust dengan menggunakan fungsi Huber akan dibandingkan dengan taksiran parameter yang didapat dengan menggunakan metode OLS berdasarkan nilai effisiensi dari taksiran yang didapat.

1.2 Perumusan Masalah

1. Bagaimana mencari taksiran parameter pada model regresi robust sederhana dengan menggunakan fungsi Huber?
2. Bagaimana mencari taksiran parameter pada model regresi robust berganda dengan menggunakan fungsi Huber?
3. Bagaimana perbandingan antara taksiran parameter model regresi robust dengan fungsi Huber dan taksiran parameter yang didapat dengan metode OLS dilihat dari efisiensi dari taksiran parameter?

1.3 Tujuan Penulisan

1. Mencari taksiran parameter pada model regresi robust sederhana dengan menggunakan fungsi Huber
2. Mencari taksiran parameter pada model regresi robust berganda dengan menggunakan fungsi Huber
3. Membandingkan antara taksiran parameter model regresi robust dengan fungsi Huber dan taksiran parameter yang didapat dengan metode OLS dilihat dari nilai effisiensi dari taksiran parameter

1.4 Pembatasan Masalah

Permasalahan dalam tulisan ini hanya dibatasi pada penaksiran parameter dan tidak dilakukan pengujian model.

1.5 Sistematika Penulisan

Bab I. Pendahuluan

- Latar belakang masalah
- Perumusan masalah
- Tujuan penulisan
- Pembatasan masalah
- Sistematika penulisan

Bab II. Landasan Teori

- Regresi inier sederhana
- Regresi linier berganda
- Identifikasi *outlier* untuk regresi linier sederhana dan regresi linier berganda
- MADN (Normalized Median Absolute Deviation)
- Taksiran Huber

Bab III. Penaksiran parameter pada model regresi robust dengan menggunakan fungsi Huber

- Penaksiran parameter pada model regresi robust sederhana dengan menggunakan fungsi Huber
- Penaksiran parameter pada model regresi robust berganda dengan menggunakan fungsi Huber
- Effisiensi taksiran parameter dengan metode regresi robust dengan menggunakan fungsi Huber dan metode OLS

Bab IV. Contoh Penerapan

- Data
- Deteksi *outlier*
- Mencari taksiran parameter dengan metode OLS yang mengikutkan data *outlier* dan dengan metode regresi robust yang juga mengikutkan *outlier*.
- Membandingkan kedua taksiran berdasarkan effisiensi dari taksiran yang diperoleh

Bab V. Penutup

- Kesimpulan
- Saran

BAB II

LANDASAN TEORI

2.1 Regresi Linier Sederhana

2.1.1 Model Regresi Linier Umum

Model regresi digunakan untuk melihat hubungan fungsional antara variabel dependen (variabel terikat) dan variabel independen (variabel bebas). Jika hubungan antara variabel independen dengan variabel dependen linier dan hanya ada satu variabel independen pada model regresi, maka model regresi disebut model regresi linier sederhana. Bentuk model regresi linier sederhana seperti berikut :

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (2.1.1)$$

Dimana : y = variabel dependen

x = variabel independen

β_j = parameter model regresi yang tidak diketahui nilainya, j=0,1, yaitu β_0 merupakan *intercept*-y dan β_1 merupakan *slope* (kemiringan) garis

ε = *random error*

Dalam analisis regresi terdapat beberapa asumsi dari *error* yang harus terpenuhi agar taksiran parameter yang diperoleh BLUE (*Best Linier Unbiased Estimator*). Asumsi-asumsi *error* antara lain

1. Komponen *error* ε mempunyai mean nol
2. Komponen *error* ε mempunyai variansi konstan σ^2 atau disebut homoskedastisitas
3. *Error* tidak berkorelasi atau $corr(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$
4. *Error* berdistribusi normal

2.1.2 Estimasi Parameter Model

Salah satu metode yang digunakan untuk menaksir parameter dalam model regresi adalah metode *Ordinary Least Square* (OLS). Prinsip dari metode OLS adalah meminimumkan jumlah kuadrat *error*. Jumlah kuadrat *error* ini dinamakan fungsi *least square* dan dinyatakan sebagai berikut :

$$\begin{aligned} S(\beta_0, \beta_1) &= \sum_{i=1}^{n} \varepsilon_i^2 \\ &= \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2 \end{aligned} \qquad (2.1.2)$$

dimana y_i = nilai variabel dependen untuk observasi ke-i

x_i = nilai variabel independen untuk observasi ke-i

Taksiran *least square* dari β_0 dan β_1 diperoleh dengan meminimumkan fungsi *least square* $S(\beta_0, \beta_1)$ terhadap β_0 dan β_1 , yaitu

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \quad (2.1.3)$$

dan

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0 \quad (2.1.4)$$

Penyederhanaan dari persamaan (2.1.3) dan (2.1.4) seperti berikut

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad (2.1.5)$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} y_i x_i \quad (2.1.6)$$

Solusi dari persamaan (2.1.5) dan (2.1.6) adalah

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.1.7)$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (2.1.8)$$

Pembuktian taksiran parameter β_0 dan β_1 dapat dilihat pada lampiran 1.

dimana $S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i \left(x_i - \bar{x}\right)$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$

2.1.3 Ekspektasi dan Variansi dari Taksiran Parameter Metode OLS

1. Nilai ekspektasi taksiran parameter dengan metode OLS

Taksiran untuk β_1 yaitu $\hat{\beta}_1$, dapat dinyatakan dalam bentuk kombinasi linier dari observasi y_i, yaitu

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n} c_i y_i \qquad (2.1.9)$$

dimana $c_i = \frac{\left(x_i - \bar{x}\right)}{S_{xx}}$ untuk i = 1,2,…,n.

Sebelum menunjukkan bahwa $\hat{\beta}_1$ merupakan taksiran yang tak bias untuk β_1, terlebih dahulu ditunjukkan bahwa $\sum_{i=1}^{n} c_i = 0$ dan $\sum_{i=1}^{n} c_i x_i = 1$.

$$\sum_{i=1}^{n} c_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i - n\bar{x}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i}{S_{xx}} = 0 \qquad (2.1.10)$$

$$\sum_{i=1}^{n} c_i x_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})x_i}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i^2 - \bar{x}n\frac{\sum_{i=1}^{n} x_i}{n}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{S_{xx}} = 1$$

(2.1.11)

Sekarang akan ditunjukkan bahwa $\hat{\beta}_1$ merupakan taksiran yang tak bias untuk β_1.

$$\begin{aligned} E\left(\hat{\beta}_1\right) &= E\left(\sum_{i=1}^{n} c_i y_i\right) \\ &= \sum_{i=1}^{n} c_i E\left(y_i\right) \\ &= \sum_{i=1}^{n} c_i \left(\beta_0 + \beta_1 x_i\right) \\ &= \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i \\ &= \beta_1 \end{aligned}$$

Hal diatas dapat dipenuhi karena $E\left(\varepsilon_i\right) = 0$ (berdasarkan asumsi *error*), dan berdasarkan (2.1.10) serta (2.1.11). Jadi, karena $E\left(\hat{\beta}_1\right) = \beta_1$, maka $\hat{\beta}_1$ merupakan taksiran yang tak bias untuk β_1.

Selanjutnya akan ditunjukkan bahwa $\hat{\beta}_0$ juga merupakan taksiran yang tak bias untuk β_0.

$$
\begin{aligned}
E\left(\hat{\beta}_0\right) &= E\left(\bar{y} - \hat{\beta}_1\bar{x}\right) \\
&= E\left(\bar{y}\right) - E\left(\hat{\beta}_1\bar{x}\right) \\
&= \frac{1}{n}E\left(\sum_{i=1}^{n} y_i\right) - \bar{x}E\left(\hat{\beta}_1\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i\right) - \bar{x}\beta_1 \\
&= \beta_0 + \beta_1 \frac{\sum_{i=1}^{n} x_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n}\beta_1 \\
&= \beta_0
\end{aligned}
$$

Jadi, karena $E\left(\hat{\beta}_0\right) = \beta_0$, maka $\hat{\beta}_0$ merupakan taksiran yang tak bias untuk β_0.

2. Variansi taksiran parameter dengan metode OLS

Variansi dari $\hat{\beta}_1$ sebagai berikut

$$
\begin{aligned}
\operatorname{var}\left(\hat{\beta}_1\right) &= \operatorname{var}\left(\sum_{i=1}^{n} c_i y_i\right) \\
&= \operatorname{var}\left(c_1 y_1 + c_2 y_2 + \ldots + c_n y_n\right) \\
&= c_1^2 \operatorname{var}\left(y_1\right) + c_2^2 \operatorname{var}\left(y_2\right) + \ldots + c_n^2 \operatorname{var}\left(y_n\right)
\end{aligned}
$$

karena $\operatorname{var}\left(y_i\right) = \sigma^2$, sehingga $\operatorname{var}\left(\hat{\beta}_1\right)$ menjadi

$$\begin{aligned}\operatorname{var}\left(\hat{\beta}_1\right) &= c_1^2\sigma^2 + c_2^2\sigma^2 + \ldots + c_n^2\sigma^2 \\ &= \sum_{i=1}^{n} c_i^2\sigma^2 \\ &= \frac{\sigma^2}{S_{xx}}\end{aligned}$$

Jadi, $\operatorname{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}}$.

Variansi dari $\hat{\beta}_0$ sebagai berikut

$$\begin{aligned}\operatorname{var}\left(\hat{\beta}_0\right) &= \operatorname{var}\left(\bar{y} - \hat{\beta}_1\bar{x}\right) \\ &= \operatorname{var}\left(\bar{y}\right) + \bar{x}^2 \operatorname{var}\left(\hat{\beta}_1\right) - 2\bar{x}\operatorname{cov}\left(\bar{y}, \hat{\beta}_1\right) \\ &= \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{S_{xx}} - 0 \\ &= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\end{aligned}$$

Jadi, $\operatorname{var}\left(\hat{\beta}_0\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$.

2.2 Regresi Linier Berganda

2.2.1 Model Regresi Linier Umum

Pada model regresi linier, dapat dijumpai variabel independen yang lebih dari satu. Kondisi seperti ini dalam model regresi disebut model regresi linier berganda. Ilustrasi data dari regresi linier berganda dengan y sebagai variabel dependen, $x_{1,} x_2, ..., x_k$ variabel independen dan n observasi, yaitu

Pengamatan	i	y	x_1	x_2	$\cdots$	x_k
	1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
	2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

Bentuk model regresi linier berganda sebagai berikut :

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \qquad , \; i = 1, 2, ..., n \end{aligned} \tag{2.2.1}$$

Model regresi linier berganda diatas dapat dijabarkan seperti :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \varepsilon_n \end{aligned} \tag{2.2.2}$$

Sebut $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

Sehingga model (2.2.2) dapat ditulis sebagai :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.2.3}$$

Dalam bentuk matriks model regresi linier berganda seperti berikut :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.2.4}$$

Dengan **y** = vektor kolom dari variabel dependen yang berukuran n x 1

X = matriks dari variabel independen yang berukuran n x p, dengan p = k + 1

β = vektor kolom dari parameter regresi yang berukuran p x 1

ε = vektor kolom dari *error* yang berukuran n x 1,

$\boldsymbol{\varepsilon} \sim NID(\mathbf{0}, \Sigma)$

2.2.2 Estimasi Parameter Model

Sama halnya dengan regresi linier sederhana, estimasi parameter dalam model regresi linier berganda juga dapat diperoleh dengan metode OLS. Fungsi *least square* untuk model regresi linier berganda sebagai berikut :

$$
\begin{aligned}
S(\beta_0, \beta_1, ..., \beta_k) &= \sum_{i=1}^{n} \varepsilon_i^2 \\
&= \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij})^2
\end{aligned} \tag{2.2.5}
$$

dimana y_i = nilai variabel dependen untuk observasi ke-i

x_{ij} = nilai observasi ke-i dari variabel x_j , dengan n > k, i=1,2,…,k

Taksiran *least square* dari $\beta_0, \beta_1, ..., \beta_k$ didapat dengan meminimumkan fungsi *least square* $S(\beta_0, \beta_1, ..., \beta_k)$ terhadap parameter-parameter pada model regresi.

$$
\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij}) = 0 \tag{2.2.6}
$$

$$
\left.\frac{\partial S}{\partial \beta_j}\right|_{\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij}) x_{ij} = 0 \quad , \ j = 1, 2, ..., k \tag{2.2.7}
$$

Persamaan (2.2.6) dan (2.2.7) diuraikan, sehingga diperoleh persamaan seperti dibawah ini

$$
n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + ... + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i \tag{2.2.8}
$$

$$
\begin{aligned}
\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} {x_{i1}}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + ... + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} &= \sum_{i=1}^{n} x_{i1} y_i \\
&\vdots \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{ik} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + ... + \hat{\beta}_k \sum_{i=1}^{n} {x_{ik}}^2 &= \sum_{i=1}^{n} x_{ik} y_i
\end{aligned} \tag{2.2.9}
$$

Dalam bentuk matriks dapat ditulis sebagai berikut :

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \varepsilon_i^{\,2} \\ &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \end{aligned} \qquad (2.2.10)$$

dimana $\boldsymbol{\varepsilon}^T$ = transpose dari $\boldsymbol{\varepsilon}$. Karena $\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}$ merupakan matriks berukuran 1 x 1 dan transpos dari $\left(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}\right)^T = \mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$ juga merupakan skalar yang sama, maka

$$S(\boldsymbol{\beta}) = \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \qquad (2.2.11)$$

$$\begin{aligned} \left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \\ \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^T\mathbf{y} \end{aligned} \qquad (2.2.12)$$

Jika matriks $\mathbf{X}^T\mathbf{X}$ *invertible*, maka diperoleh taksiran *least square* untuk β, yaitu

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad (2.2.13)$$

2.2.3 Ekspektasi dan Matriks Kovariansi dari Taksiran Parameter dengan Metode OLS

1. Nilai ekspektasi taksiran parameter dengan metode OLS

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E\left(\mathbf{y}\right) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E\left(\mathbf{X}\boldsymbol{\beta}\right) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\; E(\boldsymbol{\beta}) \\ &= \boldsymbol{\beta} \end{aligned} \tag{2.2.14}$$

karena $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ dan $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{I}$.

Jadi, karena $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, maka $\hat{\boldsymbol{\beta}}$ merupakan taksiran yang tak bias untuk $\boldsymbol{\beta}$.

2. Matriks kovariansi untuk $\hat{\boldsymbol{\beta}}$

$$\begin{aligned} \operatorname{cov}(\hat{\boldsymbol{\beta}}) &= E\left[\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)^T\right] \\ &= E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T\right] \end{aligned} \tag{2.2.15}$$

Substitusikan persamaan $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ kedalam persamaan (2.2.13), sehingga diperoleh

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} \end{aligned} \tag{2.2.16}$$

Substitusikan hasil dari persamaan (2.2.16) ke persamaan (2.2.15), didapat

$$\begin{aligned}\text{cov}(\hat{\boldsymbol{\beta}}) &= E\left[\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right)\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right)^T\right] \\ &= E\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right] \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\end{aligned}$$

Jadi, $\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$. Jika $C = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, maka $\text{var}\left(\hat{\beta}_j\right) = \sigma^2 C_{jj}$ dan $\text{cov}\left(\hat{\beta}_i, \hat{\beta}_j\right) = \sigma^2 C_{ij}$, dimana C_{jj} merupakan elemen diagonal ke-j dari $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$.

2.3 Identifikasi *Outlier* untuk Regresi Linier Sederhana dan Berganda

Outlier merupakan suatu observasi yang menyimpang jauh dari hubungan linier yang dibentuk oleh mayoritas dari data. Ada beberapa cara untuk mengidentifikasi *outlier* dalam analisis regresi, antara lain *leverage*, *scaled residuals*, DFFITS, DFBETAS, dan *Cook's Distance*.

1. *Leverage*

   Besarnya pengaruh suatu observasi terhadap besarnya taksiran parameter antara lain dapat dilihat dari jarak nilai x terhadap pusat nilai x semua observasi. Suatu observasi yang mempunyai nilai x yang jauh dari pusat nilai x dapat berpengaruh kuat dalam analisis regresi. Karena itu, nilai x yang jauh dari pusat perlu dideteksi, salah satunya dengan elemen diagonal dari *hat matriks*. Misal $x_1, x_2, ..., x_n$ variabel independen, n banyaknya observasi. Sebut $\mathbf{x}_i^T = (x_1, x_2, ..., x_k)$. *Hat matriks* didefinisikan sebagai

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \qquad (2.3.1)$$

Elemen ke-i pada diagonal dari *hat matriks*, sebut h_{ii} , dapat diperoleh dari

$$h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i \qquad (2.3.2)$$

h_{ii} menyatakan jarak dari $\mathbf{x}_i$ ke pusat nilai **x** dari semua observasi. h_{ii} disebut *leverage* dari observasi ke-i. Untuk regresi linier sederhana, h_{ii} dapat dirumuskan sebagai

$$h_{ii} = \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{S_{xx}} \qquad (2.3.3)$$

Dapat ditunjukkan bahwa $\sum_{i=1}^{n} h_{ii} = p$, dimana p merupakan banyaknya parameter dalam persamaan regresi termasuk *intercept*. Rata-rata dari h_{ii} , sebut $\bar{h} = \frac{p}{n}$. Nilai h_{ii} dikatakan besar jika nilainya lebih dari dua kali rata-rata nilai h_{ii}, yaitu

$h_{ii} > \frac{2p}{n}$ (Hoaglin&Welsch 1978). Pada regresi linier sederhana, p=2, maka observasi berpotensi sebagai *outlier* jika $h_{ii} > \frac{4}{n}$. Jika h_{ii} besar, maka jarak $\mathbf{x}_i$ terhadap pusat $\mathbf{x}$ besar, sehingga observasi ke-i merupakan *outlier*.

2. *Scaled Residual*

Scaled residual merupakan *residual* yang nilainya telah distandarkan. Ukuran yang diperoleh dari *scaled residual* ini akan terbebas dari skala, sehingga dapat dipakai untuk menentukan apakah observasi merupakan *outlier* atau bukan berdasarkan nilai *scaled residuals*. Ada beberapa ukuran dari *scaled residual* ini, antara lain *Standardized Residuals, Studentized Residuals, PRESS Residuals,* dan *R-student.*

a) *Standardized Residuals*

Variansi *residual* ditaksir dengan MSE.

Standardized Residuals dari observasi ke-i didefinisikan sebagai :

$$d_i = \frac{e_i}{\sqrt{MSE}}, i = 1, 2, ..., n \qquad (2.3.4)$$

dimana e_i = *residual* observasi ke-i

$$MSE = \frac{\sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2}{n-p}$$

Untuk regresi linier sederhana, $MSE = \dfrac{\sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2}{n-2}$.

Suatu observasi berpotensi sebagai *outlier* jika mempunyai *standardized redisuals* $|d_i| > 2$.

b) *Studentized Residuals*

Menggunakan MSE sebagai taksiran variansi dari *residual* ke-i, e_i , hanya merupakan suatu pendekatan. *Exact* variansi *residual* adalah $\operatorname{var}\left(e_i\right) = \sigma^2\left(1 - h_{ii}\right)$. Hal ini dikarenakan $\mathbf{e} = \left(\mathbf{I} - \mathbf{H}\right)\mathbf{y}$ dimana $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$,

$$
\begin{aligned}
\mathbf{e} &= \left(\mathbf{I} - \mathbf{H}\right)\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon} \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon} \\
&= \left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}
\end{aligned}
$$

Sedemikian sehingga

$$
\begin{aligned}
\operatorname{var}\left(\mathbf{e}\right) &= \operatorname{var}\left[\left(\mathbf{I} - \mathbf{H}\right)\boldsymbol{\varepsilon}\right] \\
&= \left(\mathbf{I} - \mathbf{H}\right)\operatorname{var}\left(\boldsymbol{\varepsilon}\right)\left(\mathbf{I} - \mathbf{H}\right)^T \\
&= \sigma^2\left(\mathbf{I} - \mathbf{H}\right)
\end{aligned}
$$

karena $\operatorname{var}\left(\boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{I}$ dan $\mathbf{I} - \mathbf{H}$ simetris dan *idempotent*.

Jadi, diperoleh $\operatorname{var}\left(e_i\right) = \sigma^2\left(1 - h_{ii}\right)$.

Studentized Residuals didefinisikan sebagai

$$r_i = \frac{e_i}{\sqrt{MSE\left(1 - h_{ii}\right)}}, i = 1, 2, ..., n \qquad (2.3.5)$$

Dari rumus (2.3.5), terlihat bahwa r_i berpengaruh pada h_{ii}. Jika h_{ii} besar (berpotensi sebagai outlier), maka r_i juga besar. Suatu observasi dikatakan berpotensial sebagai *outlier* jika mempunyai *studentized residuals* $|r_i| > 2$. Jadi, r_i dapat digunakan untuk mendeteksi *outlier* secara simultan.

c) *PRESS Residuals* (*deleted residuals*)

Pendekatan *PRESS residual* dalam melihat observasi yang merupakan *outlier* berbeda dengan *standardized residuals* dan *studentized residuals*. *PRESS residuals* melihat selisih antara nilai observasi ke-I, y_i dengan nilai taksiran y_i yang didapat dari model tanpa menyertakan observasi ke-i.

PRESS residuals didefinisikan sebagai :

$$e_{(i)} = y_i - \hat{y}_{(i)}, \quad i = 1, 2, ..., n \qquad (2.3.6)$$

Dimana y_i = nilai observasi ke-i

$\hat{y}_{(i)}$ = nilai taksiran yang didapat dari model dengan mengeluarkan observasi ke-i

PRESS residuals dapat juga dituliskan dalam bentuk

$$e_{(i)} = \frac{e_i}{1 - h_{ii}} \qquad (2.3.7)$$

(Montgomery, Peck&Vining 2001). Dari persamaan (2.3.7), terlihat bahwa nilai $e_{(i)}$ juga tergantung pada h_{ii}, sehingga dengan semakin besarnya h_{ii}, yang mengindikasikan bahwa observasi tersebut berpotensi sebagai *outlier*, maka nilai $e_{(i)}$ yang diperoleh juga semakin besar. Oleh karena itu, *PRESS residual*s dapat digunakan untuk mendeteksi *outlier* yang terjadi secara simultan.

d) *R-student*

Perhitungan *R-student* menyerupai *studentized residual*s, akan tetapi variansi *residual* yang digunakan untuk *R-student* memperhitungkan saat observasi ke-i dikeluarkan dari pengamatan, sehingga variansi *residual* ditaksir dengan $S_{(i)}^2$, yaitu

$$S_{(i)}^2 = \frac{(n-p)MSE - \dfrac{e_i^2}{(1-h_{ii})}}{n-p-1} \qquad (2.3.8)$$

R-student didefinisikan sebagai

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2 (1-h_{ii})}}, \quad i = 1, 2, ..., n \qquad (2.3.9)$$

Terlihat dari persamaan (2.3.9), R-student juga dapat digunakan untuk melihat indikasi *outlier* secara simultan, karena t_i juga bergantung pada nilai h_{ii}. Suatu observasi berpotensi sebagai *outlier* jika nilai $|t_i| > t_{\frac{\alpha}{2}, n-p-1}$

(Montgomery, Peck&Vining 2001).

3. DFFITS

DFFIT$_i$ (*difference in fit*) yaitu ukuran pengaruh dengan melihat selisih nilai taksiran dari observasi ke-i ($\hat{y}_i$) dengan nilai taksiran dari observasi ke-i berdasarkan model jika observasi ke-i dikeluarkan dari pengamatan ($\hat{y}_{(i)}$). Jadi, DFFIT$_i$ dapat dituliskan sebagai $DFFIT_i = \hat{y}_i - \hat{y}_{(i)}$. Semakin besar selisih antara nilai $\hat{y}_i$ dan $\hat{y}_{(i)}$, maka data ke-i semakin berpengaruh.

DFFITS$_i$ merupakan DFFIT$_i$ yang dibagi standar errornya, yaitu

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}} \qquad (2.3.10)$$

dimana $S_{(i)}^2$ = variansi sampel dari residual yang diperoleh dari model jika observasi ke-i dikeluarkan dari pengamatan

DFFITS$_i$ bergantung pada h$_{ii}$, sehingga ukuran deteksi *outlier* dengan DFFITS$_i$ memperhatikan nilai x dan y secara simultan. Belsley, Kuh, dan Welsch menyatakan bahwa suatu observasi dikatakan berpotesi sebagai *outlier* jika

$|\mathrm{DFFITS}_i| > 2\sqrt{\frac{p}{n}}$, dengan n menyatakan banyaknya observasi dan p menyatakan banyaknya parameter dalam model. Jika model merupakan regresi linier sederhana, p=2, observasi berpotesi sebagai *outlier* jika $|\mathrm{DFFITS}_i| > 2\sqrt{\frac{2}{n}}$.

4. $\mathrm{DFBETAS}_{j,i}$

DFBETA$_{j,i}$ (*difference in Beta*) yaitu ukuran pengaruh dengan melihat selisih nilai taksiran koefisien regresi ke-j, $\hat{\beta}_j$ dengan nilai taksiran koefisien regresi ke-j saat observasi ke-i dikeluarkan, $\hat{\beta}_{j(i)}$. Untuk model regresi sederhana, DFBETA$_0$ merupakan beda *intercept* dan DFBETA$_1$ merupakan beda *slope* (kemiringan) dari garis regresi.

Jika selisih $\hat{\beta}_0$ dengan $\hat{\beta}_{0(i)}$ besar , maka memperlihatkan bahwa observasi ke-i cukup mempengaruhi parameter regresi. Demikian juga dengan β_1, jika selisih $\hat{\beta}_1$ dengan $\hat{\beta}_{1(i)}$ besar (kemiringan yang cukup berbeda antara $\hat{y}_i$ dengan $\hat{y}_{(i)}$), maka observasi ke-i cukup mempengaruhi parameter regresi.

DFBETAS$_i$ merupakan DFBETA$_i$ yang dibagi dengan standar errornya, yaitu

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}} \tag{2.3.11}$$

dimana $\hat{\beta}_j$ = taksiran koefisien regresi ke-j

$\hat{\beta}_{j(i)}$ = taksiran koefisien regresi ke-j saat observasi ke-i dikeluarkan

$S_{(i)}^2$ = variansi sampel saat observasi ke-i dikeluarkan

C_{jj} = elemen diagonal ke-j dari $(X^TX)^{-1}$

Belsley, Kuh, dan Welsch menyatakan bahwa suatu observasi dikatakan berpengaruh jika $|DFBETAS_{j,i}| > \frac{2}{\sqrt{n}}$ dengan n menyatakan banyaknya observasi.

5. *Cook's Distance*

Cook's distance merupakan ukuran pengaruh observasi ke-i terhadap semua taksiran parameter regresi. *Cook's distance* didefinisikan sebagai :

$$D_i = \frac{(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X})(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})}{pMSE} = \frac{\left(y_i - \hat{y}_i\right)^2}{pMSE}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right] \quad (2.3.12)$$

dimana $\hat{\beta}$ = vektor taksiran koefisien regresi

$\hat{\beta}_{(i)}$ = vektor taksiran koefisien regresi tanpa observasi ke-i

Dari rumus diatas, *Cook's distance* bergantung pada *residual* $\left(y_i - \hat{y}_i\right)$ dan *leverage* h_{ii} untuk observasi ke-i. Pengaruh observasi diukur oleh jarak D_i. Nilai D_i besar mengindikasikan bahwa observasi ke-i berpotensi sebagai *outlier*. Observasi dengan D_i >1 sudah dapat dikatakan sebagai *outlier* (Montgomery, Peck&Vining 2001).

2.4 MADN (Normalized Median Absolute Deviation)

Misal $x_1, x_2, ..., x_n$ nilai-nilai dari sampel random dari distribusi yang mempunyai mean μ dan variansi σ^2. x_i dapat dituliskan dalam bentuk sebagai berikut :

$$x_i = \mu + u_i \qquad \left(i = 1, 2, ..., n\right) \quad (2.4.1)$$

Misal u_i mempunyai fungsi distribusi F_0 $\forall i = 1, 2, ..., n$ dan u_i saling bebas.

Didefinisikan $MAD(x) = MAD\left(x_1, x_2, ..., x_n\right)$

$$= med\left\{\left|x - med\left(x\right)\right|\right\}$$

dimana median adalah ukuran pusat data yang *robust* terhadap *outlier*.

Jika x simetris, maka $med(x) = \mu$, sehingga diperoleh $MAD(x) = med\left\{\left|x_i - \mu\right|\right\}$

dan berlaku

$$\Pr\left\{\left|x_i - \mu\right| \le MAD(x)\right\} = \frac{1}{2}$$

$$\Pr\left(-MAD(x) \le x_i - \mu \le MAD(x)\right) = \frac{1}{2}$$

$$\Pr\left(-\frac{MAD(x)}{\sigma} \le \frac{x_i - \mu}{\sigma} \le \frac{MAD(x)}{\sigma}\right) = \frac{1}{2}$$

$$\Pr\left(-\frac{MAD(x)}{\sigma} \le Z \le \frac{MAD(x)}{\sigma}\right) = \frac{1}{2}$$

Jika $Z \sim N\left(0,1\right)$, maka didapat $\dfrac{MAD(x)}{\sigma} = 0.6745$.

$\hat{\sigma} = \dfrac{MAD(x)}{0.6745}$ disebut *Normalized Median Absolute Deviation* (MADN(x)).

MADN(x) merupakan taksiran yang *robust* untuk σ.

2.5 Taksiran Huber

Taksiran Huber merupakan salah satu taksiran yang termasuk dalam *M-estimator* yang *robust* terhadap adanya *outlier*. *M-estimator* dicari berlandaskan konsep dari taksiran *maksimum likelihood* yang memaksimumkan fungsi *likelihood*.

Misal $x_1, x_2, ..., x_n$ nilai-nilai dari sampel *random* dari distribusi yang mempunyai mean μ. x_i dapat dituliskan dalam bentuk sebagai berikut :

$$x_i = \mu + u_i \qquad (i = 1, 2, ..., n) \tag{2.5.1}$$

Model diatas disebut model lokasi. Misal u_i mempunyai fungsi distribusi $F_0 \;\; \forall i = 1, 2, ..., n$ dan u_i saling bebas. Observasi $x_1, x_2, ..., x_n$ mempunyai distribusi dengan fungsi distribusi $F_0(x-\mu)$, dengan F_0 adalah fungsi distribusi dari u_i. $f_0 = F_0^{'}$ adalah pdf dari u_i.

Fungsi *likelihood* dari observasi $x_1, x_2, ..., x_n$ adalah

$$L(\mu; x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_0(x_i - \mu) \tag{2.5.2}$$

$$\begin{aligned} \log L(\mu; x_1, x_2, ..., x_n) &= \log\left(\prod_{i=1}^{n} f_0(x_i - \mu)\right) \\ &= \sum_{i=1}^{n} \log f_0(x_i - \mu) \end{aligned} \tag{2.5.3}$$

MLE dari μ adalah nilai μ yang memaksimumkan $L(\mu; x_1, x_2, ..., x_n)$ atau nilai μ yang memaksimumkan $\log L(\mu; x_1, x_2, ..., x_n)$.

Sebut $\hat{\mu} = \arg\max_{\mu} L(\mu; x_1, x_2, ..., x_n)$

$$= \arg\max_{\mu} \log L(\mu; x_1, x_2, ..., x_n) \qquad (2.5.4)$$

Misal $\rho = -\log f_0$, sehingga $-\log f_0(x_i - \mu) = \rho(x_i - \mu)$. Oleh karena itu, (2.5.4) dapat dinyatakan sebagai :

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} \rho(x_i - \mu) \qquad (2.5.5)$$

Perhatikan 2 kasus dibawah ini :

1. Jika F_0 = N(0,1), maka

$$f_0(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \qquad (2.5.6)$$

Karena $\rho = -\log f_0$, maka diperoleh $\rho(u) = \frac{u^2}{2} - \log\frac{1}{\sqrt{2\pi}}$. Jadi, $\rho(x_i - \mu)$ dapat dituliskan seperti berikut :

$$\rho(x_i - \mu) = \frac{(x_i - \mu)^2}{2} - \log\frac{1}{\sqrt{2\pi}} \qquad (2.5.7)$$

Persamaan (2.5.5) menjadi

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} \left\{ \frac{(x_i - \mu)^2}{2} - \log \frac{1}{\sqrt{2\pi}} \right\}$$
$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2} \qquad (2.5.8)$$

Untuk mendapatkan nilai μ yang meminimumkan $\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2}$, maka $\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0$ dimana $\psi = \rho'$. Jika $\rho(u) = \frac{u^2}{2}$, maka $\psi(u) = u$, sehingga

$$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = \sum_{i=1}^{n} (x_i - \hat{\mu}) = 0 \qquad (2.5.9)$$

Persamaan (2.5.9) dijabarkan untuk mendapatkan nilai $\hat{\mu}$, yaitu

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \hat{\mu} = 0$$
$$n\hat{\mu} = \sum_{i=1}^{n} x_i$$
$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$\hat{\mu} = \bar{x}$$

Jadi, jika F_0 = N(0,1) maka MLE dari μ adalah $\bar{x}$ (mean sampel). $\bar{x}$ tidak *robust* terhadap *outlier*.

2. Jika F_0 = *double exponential*, maka

$$f_0(u) = \frac{1}{2} e^{-|u|} \tag{2.5.10}$$

Karena $\rho = -\log f_0$, maka diperoleh $\rho(u) = |u| - \log \frac{1}{2}$, sehingga

$$\rho(x_i - \mu) = |x_i - \mu| - \log \frac{1}{2} \tag{2.5.11}$$

Persamaan (2.5.5) menjadi

$$\begin{aligned} \hat{\mu} &= \arg\min_{\mu} \sum_{i=1}^{n} \left\{ |x_i - \mu| - \log \frac{1}{2} \right\} \\ \hat{\mu} &= \arg\min_{\mu} \sum_{i=1}^{n} |x_i - \mu| \end{aligned} \tag{2.5.12}$$

Untuk mendapatkan nilai μ yang meminimumkan $\sum_{i=1}^{n} |x_i - \mu|$, maka $\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0$ dimana $\psi = \rho'$. Karena $\rho(u) = |u|$ tidak dapat differensial di titik u = 0, maka didefinisikan fungsi

$$\psi(u) = sign(u) = \begin{cases} -1 & \text{jika u<0} \\ 0 & \text{jika u=0} \\ 1 & \text{jika u>0} \end{cases} \tag{2.5.13}$$

Definisikan $I(u > 0) = \begin{cases} 1 & \text{jika u>0} \\ 0 & \text{jika u} \leq 0 \end{cases}$ dan $I(u < 0) = \begin{cases} 1 & \text{jika u<0} \\ 0 & \text{jika u} \geq 0 \end{cases}$

Fungsi $sign(u)$ dapat dinyatakan sebagai

$$sign(u) = I(u > 0) - I(u < 0) \tag{2.5.14}$$

Dengan mensubstitusikan persamaan (2.5.14) ke

$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0$ diperoleh

$$\begin{aligned} \sum_{i=1}^{n} sign(x_i - \hat{\mu}) &= 0 \\ \sum_{i=1}^{n} \left(I(x_i - \hat{\mu} > 0) - I(x_i - \hat{\mu} < 0) \right) &= 0 \\ \sum_{i=1}^{n} I(x_i - \hat{\mu} > 0) - \sum_{i=1}^{n} I(x_i - \hat{\mu} < 0) &= 0 \\ \#(x_i > \hat{\mu}) - \#(x_i < \hat{\mu}) &= 0 \end{aligned} \tag{2.5.15}$$

dimana $\#(x)$ = banyaknya x, sehingga $\#(x_i > \hat{\mu}) = \#(x_i < \hat{\mu})$. Hal ini menyatakan bahwa $\hat{\mu}$ adalah median. Jadi, jika F_0 = *double exponential*, maka MLE dari μ adalah median (x). Median *robust* terhadap *outlier*.

Dari penjelasan diatas dapat disimpulkan bahwa bila $\rho(x-\mu) = (x-\mu)^2$ maka MLE dari μ adalah $\bar{x}$, sedangkan bila $\rho(x-\mu) = |x-\mu|$ maka MLE dari μ adalah median(x).

Untuk selanjutnya MLE dari μ akan dicari berdasarkan pada fungsi $\rho(u)$.

Huber mendefinisikan fungsi Huber sebagai berikut :

$$\rho(u)=\begin{cases}\frac{1}{2}u^2 & \text{jika } |\text{u}|\le k\\ k|u|-\frac{1}{2}k^2 & \text{jika } |\text{u}|>k\end{cases} \tag{2.5.16}$$

Taksiran Huber untuk mean adalah nilai μ yang meminimumkan

$\sum_{i=1}^{n}\rho(x_i-\mu)$, sebut $\hat{\mu}_{Huber}$.

$$\hat{\mu}_{Huber}=\arg\min_{\mu}\sum_{i=1}^{n}\rho(x_i-\mu)$$

Karena $\hat{\mu}_{Huber}$ meminimumkan $\sum_{i=1}^{n}\rho(x_i-\mu)$, maka $\sum_{i=1}^{n}\psi(x_i-\mu)=0$, dimana

$\psi=\rho'$. Berdasarkan fungsi Huber (2.5.16), dapat diperoleh $\psi(u)$ seperti :

$$\psi(u)=\begin{cases}\text{u} & \text{jika } |\text{u}|\le k\\ k.sign(u) & \text{jika } |\text{u}|>k\end{cases} \tag{2.5.17}$$

- Untuk nilai $k\to\infty$, fungsi Huber $\rho(u_i)=\frac{1}{2}u_i^2 \quad \forall\text{i=1,2,...,n}$

  Nilai taksiran untuk mean akan diperoleh dengan

$$\begin{aligned}\hat{\mu}_{Huber}&=\arg\min_{\mu}\sum_{i=1}^{n}\rho(x_i-\mu)\\ &=\arg\min_{\mu}\sum_{i=1}^{n}\frac{1}{2}(x_i-\mu)^2\end{aligned}$$

$$\sum_{i=1}^{n}(x_i-\hat{\mu}_{Huber})=0$$

$$\hat{\mu}_{Huber}=\bar{x}$$

Jadi, taksiran Huber untuk mean saat $k \to \infty$ adalah $\bar{x}$.

- Untuk nilai $k \to 0$, fungsi Huber $\rho(u_i) = k|u_i| - \frac{1}{2}k^2$ $\forall$i=1,2,...,n

Nilai taksiran untuk mean akan diperoleh dengan

$$\begin{aligned}\hat{\mu}_{Huber} &= \arg\min_{\mu} \sum_{i=1}^{n} \left\{ k|x_i - \mu| - \frac{1}{2}k^2 \right\} \\ &= \arg\min_{\mu} \sum_{i=1}^{n} \left\{ k|x_i - \mu| \right\}\end{aligned}$$

Untuk mendapatkan nilai μ yang meminimumkan $\sum_{i=1}^{n} k|x_i - \mu|$, maka

$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0$ dimana $\psi = \rho'$ dan

$$\psi(u) = k.sign(u) = \begin{cases} -k & \text{jika u<0} \\ k & \text{jika u>0} \end{cases}$$

Dengan mensubstitusikan persamaan (2.5.14) ke $\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0$

diperoleh

$$\begin{aligned}\sum_{i=1}^{n} k.sign(x_i - \hat{\mu}_{Huber}) &= 0 \\ k\sum_{i=1}^{n} \left(I(x_i - \hat{\mu}_{Huber} > 0) - I(x_i - \hat{\mu}_{Huber} < 0) \right) &= 0 \\ k\left(\sum_{i=1}^{n} I(x_i - \hat{\mu}_{Huber} > 0) - \sum_{i=1}^{n} I(x_i - \hat{\mu}_{Huber} < 0) \right) &= 0 \\ k\left(\#(x_i > \hat{\mu}_{Huber}) - \#(x_i < \hat{\mu}_{Huber}) \right) &= 0\end{aligned}$$

sehingga $\#\left(x_i > \hat{\mu}_{Huber}\right) = \#\left(x_i < \hat{\mu}_{Huber}\right)$. Hal ini menyatakan bahwa $\hat{\mu}_{Huber}$ adalah median.

Jadi, taksiran Huber untuk mean saat $k \to 0$ adalah $med(x)$.

Jika F_0 = N(0,1), dengan simulasi data, Huber mendapatkan bahwa diperoleh effisiensi relatif taksiran Huber dengan k = 1.345 terhadap taksiran Huber dengan $k \to \infty$ adalah 95%, sehingga fungsi Huber sering ditulis sebagai :

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{jika } |u| \le 1.345 \\ 1.345|u| - \frac{1}{2}(1.345)^2 & \text{jika } |u| > 1.345 \end{cases}$$

Definisi 1. Fungsi Pengaruh

Pandang $\psi(u) = \dfrac{\partial \rho(u)}{\partial u}$. Fungsi $\psi(u)$ mengukur pengaruh dari sebuah data terhadap taksiran parameter. $\psi(u)$ disebut fungsi pengaruh (*influence function*).

Jika $\rho(u) = \dfrac{u^2}{2}$, maka $\psi(u) = u$. Pengaruh taksiran suatu data terhadap taksiran parameter secara linier sejalan dengan naiknya u.

Sedangkan jika $\rho(u) = |u|$, maka $\psi(u) = sign(u) = \begin{cases} -1 & \text{jika u<0} \\ 0 & \text{jika u=0} \\ 1 & \text{jika u>0} \end{cases}$.

Pengaruh suatu data terhadap taksiran terbatas antara [-1,1].

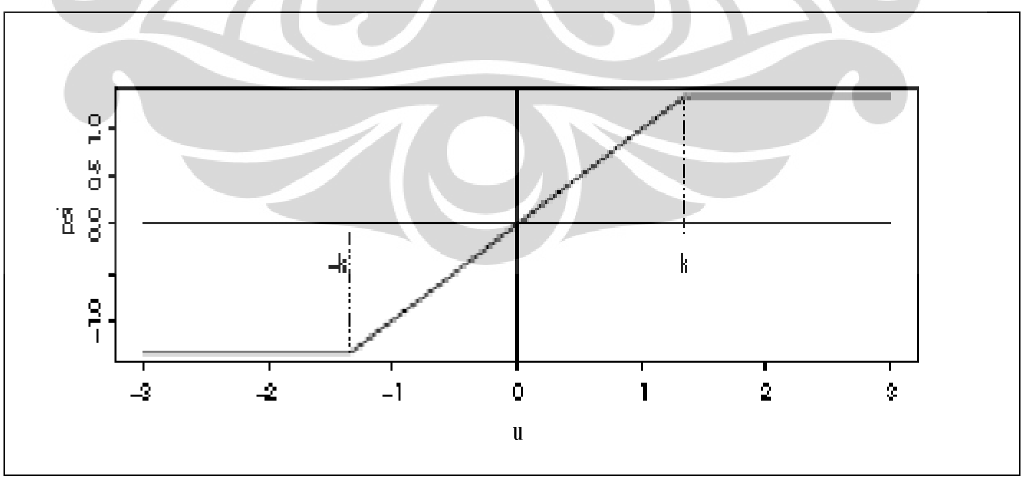
Seperti yang telah didefinisikan sebelumnya, fungsi Huber secara umum didefinisikan sebagai berikut :

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{jika } |\mathrm{u}| \le k \\ k|u| - \frac{1}{2}k^2 & \text{jika } |\mathrm{u}| > k \end{cases}$$

dan fungsi pengaruh Huber yaitu

$$\psi(u) = \begin{cases} \mathrm{u} & \text{jika } |\mathrm{u}| \le k \\ k.sign(u) & \text{jika } |\mathrm{u}| > k \end{cases}$$

Berikut ditampilkan plot dari fungsi pengaruh Huber



Gambar 1. Plot fungsi pengaruh Huber

(Sumber : Robust Statistics Theory and methods, 2006:26)

Jika k *finite*, maka $\psi(u)$ akan terbatas, sehingga pengaruh suatu data terhadap taksiran terbatas.

Definisi 2. Fungsi Bobot

Misal $\rho(u) = -\log f_0(u)$ dan $\psi(u)$ adalah fungsi pengaruh. Didefinisikan fungsi bobot :

$$w(u) = \begin{cases} \dfrac{\psi(u)}{u} & \text{jika } u \neq 0 \\ \psi'(0) & \text{jika } u=0 \end{cases} \tag{2.5.18}$$

Untuk mendapatkan nilai μ yang meminimumkan $\sum_{i=1}^{n} \rho(x_i - \mu)$, maka

$$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0 \tag{2.5.19}$$

Sehingga berdasarkan (2.5.18), persamaan (2.5.19) dapat ditulis

$$\sum_{i=1}^{n} (x_i - \hat{\mu}) w(x_i - \hat{\mu}) = 0 \quad \text{untuk } x_i \neq 0 \tag{2.5.20}$$

dan dapat diperoleh

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \text{ dengan } w_i = w(x_i - \hat{\mu}) \tag{2.5.21}$$

Karena w_i bergantung pada $\hat{\mu}$, maka $\hat{\mu}$ dicari dengan iterasi sebagai berikut :

- Pilih taksiran awal $\hat{\sigma}_0 = MADN = \dfrac{med\left\{\left|x - med\left(x\right)\right|\right\}}{0.6745}$ dan $\hat{\mu}_0 = median(x)$,

  sehingga dapat dihitung $w_{0,i} = w\left(\dfrac{x_i - \hat{\mu}_0}{\hat{\sigma}_0}\right)$.

- Pada setiap iterasi ke-t, hitung bobot $w_{t,i} = w\left(\dfrac{x_i - \hat{\mu}_t}{\hat{\sigma}_t}\right)$ dari iterasi

  sebelumnya.

- Hitung taksiran mean terboboti yang baru

$$\hat{\mu}_{t+1} = \frac{\sum_{i=1}^{n} w_{t,i} x_i}{\sum_{i=1}^{n} w_{t,i}}$$

- Iterasi berhenti jika $\left|\hat{\mu}_{t+1} - \hat{\mu}_t\right| < \varepsilon; \varepsilon > 0$.

  $\hat{\mu}$ menyatakan estimator dari mean terboboti .

BAB III

PENAKSIRAN PARAMETER PADA MODEL REGRESI ROBUST DENGAN MENGGUNAKAN FUNGSI HUBER

3.1 Penaksiran Parameter pada Model Regresi Robust Sederhana dengan Menggunakan Fungsi Huber

Misalkan hubungan linier antara satu variabel dependen dan satu variabel independen dapat dimodelkan sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (3.1.1)$$

Dimana y_i = variabel dependen observasi ke-i

x_i = variabel independen observasi ke-i

ε_i = *random error* ke-i

Seperti yang telah dijelaskan dalam bab sebelumnya, biasanya parameter regresi β ditaksir dengan menggunakan metode OLS, tetapi taksiran yang diperoleh sangat dipengaruhi oleh adanya *outlier*, maka dalam bagian ini akan dibahas mengenai taksiran parameter dalam model regresi yang *robust* terhadap *outlier*.

Metode yang paling umum untuk menaksir parameter pada regresi robust adalah dengan menggunakan *M-estimator*. Metode *M-estimator* akan mengganti kuadrat error, ε_i^2 dengan fungsi dari error.

Taksiran parameter β_0 dan β_1 diperoleh dengan cara meminimumkan fungsi $\sum_{i=1}^{n} \rho(\varepsilon_i)$ terhadap β_0 dan β_1, yaitu :

$$\frac{\partial \sum_{i=1}^{n} \rho\left(\varepsilon_i\right)}{\partial \beta_0} = 0 \qquad (3.1.2)$$

dan

$$\frac{\partial \sum_{i=1}^{n} \rho\left(\varepsilon_i\right)}{\partial \beta_1} = 0 \qquad (3.1.3)$$

Sehingga diperoleh

$$\sum_{i=1}^{n} \psi\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \qquad (3.1.4)$$

dan

$$\sum_{i=1}^{n} x_i \psi\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \qquad (3.1.5)$$

dimana ψ = turunan pertama dari ρ

Sebut $e_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$ dan didefinisikan suatu fungsi pembobot $w_i = \frac{\psi(e_i)}{e_i} = \frac{\psi\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)}{y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i}$, sehingga diperoleh $\psi\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = w_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)$. Oleh karena itu, persamaan (3.1.4) dan (3.1.5) dapat dinyatakan sebagai berikut :

$$\sum_{i=1}^{n} w_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \tag{3.1.6}$$

dan

$$\sum_{i=1}^{n} x_i w_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \tag{3.1.7}$$

Persamaan (3.1.6) dapat diselesaikan dan diperoleh taksiran untuk β_0 seperti berikut :

$$\sum_{i=1}^{n} y_i w_i - \hat{\beta}_0 \sum_{i=1}^{n} w_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i w_i = 0$$

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} y_i w_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i w_i$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \tag{3.1.8}$$

Untuk mendapatkan taksiran β_1, persamaan (3.1.7) dapat diselesaikan

$$\sum_{i=1}^{n} x_i w_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} x_i w_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i w_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i w_i x_i = 0 \qquad (3.1.9)$$

Substitusikan (3.1.8) pada persamaan (3.1.9)

$$\sum_{i=1}^{n} x_i w_i y_i - \left(\frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \right) \sum_{i=1}^{n} x_i w_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 w_i = 0$$

$$\sum_{i=1}^{n} x_i w_i y_i - \frac{\sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} + \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i w_i \sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 w_i = 0 \qquad (3.1.10)$$

Persamaan (3.1.10) dikalikan dengan $\sum_{i=1}^{n} w_i$ pada kedua sisi, sehingga

diperoleh

$$\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i w_i y_i - \sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i + \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i w_i \right)^2 - \hat{\beta}_1 \sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i^2 w_i = 0$$

$$\hat{\beta}_1 \sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i^2 w_i - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i w_i \right)^2 = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i w_i y_i - \sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i$$

$$\hat{\beta}_1 \left(\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i^2 w_i - \left(\sum_{i=1}^{n} x_i w_i \right)^2 \right) = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i w_i y_i - \sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i w_i y_i - \sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i^2 w_i - \left(\sum_{i=1}^{n} x_i w_i \right)^2} \qquad (3.1.11)$$

dimana w_i = bobot dari observasi ke-i.

Jika fungsi ψ tidak linier, maka persamaan diselesaikan dengan metode iterasi. Metode yang digunakan adalah iterasi kuadrat terkecil terboboti (*iteratively reweighted least square*) dengan prosedur :

1. Pilih nilai taksiran awal $\hat{\beta}_0^{(0)}$ dan $\hat{\beta}_1^{(0)}$, sehingga dapat dihitung

$$w_i^{(0)} = \frac{\psi\left(y_i - \hat{\beta}_0^{(0)} - \hat{\beta}_1^{(0)} x_i \right)}{y_i - \hat{\beta}_0^{(0)} - \hat{\beta}_1^{(0)} x_i}$$

2. Pada setiap iterasi ke-t, hitung *residual* $e_i^{(t-1)} = y_i - \hat{\beta}_0^{(t-1)} - \hat{\beta}_1^{(t-1)} x_i$ dan bobot $w_i^{(t-1)} = w\left[e_i^{(t-1)} \right]$ dari iterasi sebelumnya.

3. Hitung taksiran kuadrat terkecil terboboti yang baru

$$\hat{\beta}_0^{(t)} = \frac{\sum_{i=1}^{n} y_i w_i^{(t-1)}}{\sum_{i=1}^{n} w_i^{(t-1)}} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i w_i^{(t-1)}}{\sum_{i=1}^{n} w_i^{(t-1)}}$$

$$\hat{\beta}_1^{(t)} = \frac{\sum_{i=1}^{n} w_i^{(t-1)} \sum_{i=1}^{n} x_i w_i^{(t-1)} y_i - \sum_{i=1}^{n} y_i w_i^{(t-1)} \sum_{i=1}^{n} x_i w_i^{(t-1)}}{\sum_{i=1}^{n} w_i^{(t-1)} \sum_{i=1}^{n} x_i^2 w_i^{(t-1)} - \left(\sum_{i=1}^{n} x_i w_i^{(t-1)} \right)^2}$$

dimana $w_i^{(t-1)}$ = bobot pada iterasi ke-(t-1)

Langkah ke-2 dan ke-3 berulang hingga taksiran parameter yang diperoleh konvergen. Dengan perkataan lain, jika $\left|\hat{\beta}_j^{(t)} - \hat{\beta}_j^{(t-1)}\right|$ cukup kecil untuk j=0,1.

Jika penaksiran parameter pada model regresi robust dicari dengan menggunakan fungsi Huber, maka bobot w dapat dicari sebagai berikut :

Pandang fungsi Huber

$$\rho(\varepsilon) = \begin{cases} \frac{1}{2}(\varepsilon)^2 & \text{jika } |\varepsilon| \le k \\ |\varepsilon| k - \frac{1}{2}k^2 & \text{jika } |\varepsilon| > k \end{cases} \tag{3.1.12}$$

Turunan dari fungsi Huber yaitu :

$$\psi(\varepsilon) = \begin{cases} -k & \text{jika } \varepsilon < -k \\ \varepsilon & \text{jika } |\varepsilon| \le k \\ k & \text{jika } \varepsilon > k \end{cases} \tag{3.1.13}$$

Sebut $y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = e_i$, maka diperoleh

$$\psi(e) = \psi\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \begin{cases} -k & \text{jika } e < -k \\ e & \text{jika } |e| \le k \\ k & \text{jika } e > k \end{cases} \tag{3.1.14}$$

Seperti telah diketahui $w_i = \frac{\psi(e_i)}{e_i}$, maka didapat

$$w_i = \begin{cases} -\dfrac{k}{e_i} & \text{jika } e_i < -k \\ 1 & \text{jika } |e_i| \le k \\ \dfrac{k}{e_i} & \text{jika } e_i > k \end{cases} \tag{3.1.15}$$

Dengan perkataan lain

$$w_i = \begin{cases} 1 & \text{jika } |e_i| \le k \\ \left|\dfrac{k}{e_i}\right| & \text{jika } |e_i| > k \end{cases} \tag{3.1.16}$$

Jika diambil nilai terstandardisasi dari e, maka berdasarkan simulasi yang dilakukan oleh Huber dipilih nilai k = 1.345, sehingga bobot menjadi seperti berikut :

$$w_i = \begin{cases} 1 & \text{jika } |e_i| \le 1.345 \\ \dfrac{1.345}{|e_i|} & \text{jika } |e_i| > 1.345 \end{cases}$$

Dengan metode iterasi kuadrat terkecil terboboti seperti yang telah dijelaskan diatas, maka taksiran $\hat{\beta}$ dapat diperoleh.

3.2 Penaksiran Parameter pada Model Regresi Robust Berganda dengan Menggunakan Fungsi Huber

Pada bagian ini akan dibahas mengenai regresi robust berganda yang melibatkan lebih dari satu variabel independen. Misalkan hubungan linier antara satu variabel dependen dengan variabel-variabel independennya dapat dimodelkan sebagai :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i, \qquad i=1,2,...,n \tag{3.2.1}$$

dimana y_i = variabel dependen observasi ke-i

x_{ij} = variabel independen ke-j dari observasi ke-i

ε_i = *random error* ke-i

β_j = parameter regresi yang tidak diketahui nilainya, j=0,1,2,…,k

Penjabaran dari model (3.2.1) dapat ditulis seperti berikut :

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \varepsilon_n \end{aligned} \tag{3.2.2}$$

Dalam bentuk matriks, model diatas dapat dituliskan sebagai berikut :

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \tag{3.2.3}$$

dimana $\mathbf{x}_i^T = [1 \quad x_{i1} \quad \cdots \quad x_{ik}]$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Taksiran parameter $\boldsymbol{\beta}$ diperoleh dengan cara meminimumkan fungsi $\sum_{i=1}^{n} \rho(\varepsilon_i)$

terhadap $\boldsymbol{\beta}$, yaitu

$$\frac{\partial \sum_{i=1}^{n} \rho(\varepsilon_i)}{\partial \boldsymbol{\beta}} = \frac{\sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \qquad (3.2.4)$$

Sehingga diperoleh

$$\sum_{i=1}^{n} x_{ij} \psi\left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right) = 0, \qquad j = 0, 1, 2, ..., k \qquad (3.2.5)$$

Dimana $\psi = \rho'$

x_{ij} = nilai observasi ke-i pada variabel independen ke-j

$x_{i0} = 1$

Sebut $e_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ dan didefinisikan suatu fungsi pembobot $w_i = \dfrac{\psi(e_i)}{e_i}$

, sehingga diperoleh $\psi\left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right) = w_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right)$, maka persamaan (3.2.5)

menjadi

$$\sum_{i=1}^{n} x_{ij} w_i \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\right) = 0, \qquad j = 0,1,2,\ldots,k \tag{3.2.6}$$

$$\sum_{i=1}^{n} x_{ij} w_i y_i - \sum_{i=1}^{n} x_{ij} w_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} = 0$$

$$\sum_{i=1}^{n} x_{ij} w_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} x_{ij} w_i y_i \tag{3.2.7}$$

Untuk j=0 dari persamaan (3.2.7), diperoleh

$$\sum_{i=1}^{n} x_{i0} w_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} x_{i0} w_i y_i$$

$$\sum_{i=1}^{n} w_i \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik}\right) = \sum_{i=1}^{n} w_i y_i$$

$$w_1\left(\hat{\beta}_0 + \hat{\beta}_1 x_{11} + \ldots + \hat{\beta}_k x_{1k}\right) + \ldots + w_n\left(\hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \ldots + x_{nk}\hat{\beta}_k\right) = w_1 y_1 + w_2 y_2 + \ldots + w_n y_n$$

$$\left(\hat{\beta}_0 w_1 + \hat{\beta}_0 w_2 + \ldots + \hat{\beta}_0 w_n\right) + \ldots + \left(\hat{\beta}_k w_1 x_{1k} + \hat{\beta}_k w_2 x_{2k} + \ldots + \hat{\beta}_k w_n x_{nk}\right) = w_1 y_1 + w_2 y_2 + \ldots + w_n y_n$$

$$\hat{\beta}_0 \sum_{i=1}^{n} w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} w_i = w_1 y_1 + w_2 y_2 + \ldots + w_n y_n \tag{3.2.8}$$

Untuk j=1 dari persamaan (3.2.7), diperoleh

$$\sum_{i=1}^{n} x_{i1} w_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} x_{i1} w_i y_i$$

$$\sum_{i=1}^{n} x_{i1} w_i \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik} \right) = \sum_{i=1}^{n} x_{i1} w_i y_i$$

$$x_{11} w_1 \left(\hat{\beta}_0 + \hat{\beta}_1 x_{11} + \ldots + \hat{\beta}_k x_{1k} \right) + \ldots + x_{n1} w_n \left(\hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \ldots + x_{nk} \right) = x_{11} w_1 y_1 + \ldots + x_{n1} w_n y_n$$

$$\left(\hat{\beta}_0 x_{11} w_1 + \ldots + \hat{\beta}_0 x_{n1} w_n \right) + \ldots + \left(\hat{\beta}_k x_{11} w_1 x_{1k} + \ldots + \hat{\beta}_k x_{n1} w_n x_{nk} \hat{\beta}_k \right) = x_{11} w_1 y_1 + \ldots + x_{n1} w_n y_n$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} w_i = x_{11} w_1 y_1 + x_{21} w_2 y_2 + \ldots + x_{n1} w_n y_n \qquad (3.2.9)$$

Demikian seterusnya, hingga j=k. Untuk j=k, diperoleh

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 w_i = x_{1k} w_1 y_1 + x_{2k} w_2 y_2 + \ldots + x_{nk} w_n y_n \qquad (3.2.10)$$

Penjabaran dari persamaan (3.2.7) untuk j=0,1,…,k dapat ditulis dalam bentuk matriks sebagai berikut :

$$\begin{bmatrix} \hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} w_i \\ \hat{\beta}_0 \sum_{i=1}^{n} x_{i1} w_i + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} w_i \\ \vdots \\ \hat{\beta}_0 \sum_{i=1}^{n} x_{ik} w_i + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} x_{ik} w_i + \ldots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 w_i \end{bmatrix} = \begin{bmatrix} w_1 y_1 + w_2 y_2 + \ldots + w_n y_n \\ x_{11} w_1 y_1 + x_{21} w_2 y_2 + \ldots + x_{n1} w_n y_n \\ \vdots \\ x_{1k} w_1 y_1 + x_{2k} w_2 y_2 + \ldots + x_{nk} w_n y_n \end{bmatrix}$$

$$\begin{bmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} x_{i1} w_i & \cdots & \sum_{i=1}^{n} x_{ik} w_i \\ \sum_{i=1}^{n} x_{i1} w_i & \sum_{i=1}^{n} x_{i1}^2 w_i & \cdots & \sum_{i=1}^{n} x_{i1} x_{ik} w_i \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik} w_i & \sum_{i=1}^{n} x_{i1} x_{ik} w_i & \cdots & \sum_{i=1}^{n} x_{ik}^2 w_i \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ x_{11} w_1 & x_{21} w_2 & \cdots & x_{n1} w_n \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} w_1 & x_{2k} w_2 & \cdots & x_{nk} w_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ x_{11} w_1 & x_{21} w_2 & \cdots & x_{n1} w_n \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} w_1 & x_{2k} w_2 & \cdots & x_{nk} w_n \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \mathbf{X}^T \mathbf{W} \mathbf{y}$$

Jadi, persamaan (3.2.7) dapat dituliskan dalam bentuk matriks, yaitu

$$\begin{aligned} \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{W} \mathbf{y} \\ \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}^T \mathbf{W} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \end{aligned} \qquad (3.2.11)$$

Dimana **W** = matriks diagonal berukuran n x n dengan elemen diagonalnya berupa bobot $w_1, w_2, \ldots, w_n$.

Jika fungsi ψ tidak linier, maka persamaan diselesaikan dengan metode iterasi. Prosedur dari metode iterasi kuadrat terkecil terboboti (*iteratively reweighted least square*) sama dengan regresi robust sederhana, yaitu :

1. Pilih nilai taksiran awal $\hat{\boldsymbol{\beta}}^{(0)}$, sehingga dapat dihitung

$$w_i^{(0)} = \frac{\psi\left(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}^{(0)}\right)}{y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}^{(0)}}$$

2. Pada setiap iterasi ke-t, hitung *residual* $e_i^{(t-1)} = y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}^{(t-1)}$ dan bobot $w_i^{(t-1)} = w\left[e_i^{(t-1)}\right]$ dari iterasi sebelumnya.

3. Hitung taksiran kuadrat terkecil terboboti yang baru

$$\hat{\boldsymbol{\beta}}^{(t)} = \left(\mathbf{X}^T\mathbf{W}^{(t-1)}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}^{(t-1)}\mathbf{y}$$

Dimana **X** matriks berukuran n x p dengan $\mathbf{x}_i^T$ baris ke-i dari **X** dan $\mathbf{W}^{(t-1)}$ adalah matriks bobot pada iterasi ke (t-1).

Langkah ke-2 dan ke-3 berulang hingga taksiran parameter yang diperoleh konvergen. Dengan perkataan lain, jika $\left|\hat{\beta}_j^{(t)} - \hat{\beta}_j^{(t-1)}\right|$ cukup kecil untuk j=0,1,…,k.

Seperti pada model regresi robust sederhana, matriks bobot $\mathbf{W}$ diperoleh dari fungsi Huber. Jika diambil nilai terstandardisasi dari e, maka berdasarkan simulasi yang dilakukan oleh Huber, dipilih nilai k =1.345, sehingga diperoleh w_i sebagai berikut :

$$w_i = \begin{cases} 1 & \text{jika } |e_i| \le 1.345 \\ \dfrac{1.345}{|e_i|} & \text{jika } |e_i| > 1.345 \end{cases}$$

dengan w_i merupakan elemen diagonal ke-i dari matriks bobot $\mathbf{W}$.

Dengan metode iterasi kuadrat terkecil terboboti seperti yang telah dijelaskan diatas, maka taksiran $\hat{\beta}$ dapat diperoleh.

3.3 Effisiensi Taksiran Parameter pada Model Regresi Robust dengan Menggunakan Fungsi Huber dan Metode OLS

Effisiensi relatif dari dua penaksir ialah rasio dari variansi kedua penaksir tersebut. Misal $\hat{\beta}_{jHuber}$ merupakan taksiran parameter regresi ke-j yang didapat dengan metode regresi robust dengan fungsi Huber dan $\hat{\beta}_{jLS}$ merupakan taksiran parameter regresi ke-j yang didapat dengan metode *Ordinary Least Square*, dimana j=0,1,…,k. Efisiensi relatif $\hat{\beta}_{jHuber}$ terhadap $\hat{\beta}_{jLS}$ dapat diukur dengan

$$eff\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) = \frac{\operatorname{var}\left(\hat{\beta}_{jHuber}\right)}{\operatorname{var}\left(\hat{\beta}_{jLS}\right)}$$

Penaksir $\hat{\beta}_{jHuber}$ dikatakan lebih efisien dibanding $\hat{\beta}_{jLS}$ jika *eff* $\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) < 1$

atau $\text{var}\left(\hat{\beta}_{jHuber}\right) < \text{var}\left(\hat{\beta}_{jLS}\right)$.

- Metode *Ordinary Least Square*

  Seperti yang telah ditunjukkan pada bab2.2, nilai variansi dari taksiran parameter pada metode *least square*, $\hat{\beta}_{jLS}$ sebagai berikut :

  $$\text{var}\left(\hat{\beta}_{jLS}\right) = \sigma^2 C_{jj}$$

  dimana C_{jj} = elemen diagonal ke-j dari matriks $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$.

  Karena $\hat{\sigma}^2 = MSE$ adalah taksiran tak bias untuk σ^2, maka $\text{var}\left(\hat{\beta}_{jLS}\right)$ dapat ditaksir dengan $\widehat{\text{var}}\left(\hat{\beta}_{jLS}\right) = \hat{\sigma}^2 C_{jj}$.

- Metode regresi robust dengan menggunakan fungsi Huber

  Seperti telah dijelaskan pada bab3.2, taksiran *robust* untuk $\hat{\boldsymbol{\beta}}$ adalah $\hat{\boldsymbol{\beta}}_{Huber} = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$.

  Menurut Maronna, Martin, dan Yohai (2006), dapat dibuktikan bahwa

  $$\text{cov}\left(\hat{\boldsymbol{\beta}}_{Huber}\right) = v\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

$$\operatorname{var}\left(\hat{\beta}_{jHuber}\right) = v\ C_{jj}$$

dimana C_{jj} = elemen diagonal ke-j dari matriks $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$

$$v = \sigma^2 \frac{E\psi\left(\frac{e_i}{\sigma}\right)^2}{\left(E\psi'\left(\frac{e_i}{\sigma}\right)\right)^2},$$ dan nilai taksiran dari v adalah

$$\hat{v} = \hat{\sigma}^2 \frac{ave_i\left(\psi\left(\frac{e_i}{\hat{\sigma}}\right)^2\right)}{\left(ave_i\psi'\left(\frac{e_i}{\hat{\sigma}}\right)\right)^2}\frac{n}{n-p}$$

$$\psi\left(\frac{e_i}{\hat{\sigma}}\right) = \begin{cases} \dfrac{e_i}{\hat{\sigma}} & ,\ \text{jika } \left|\dfrac{e_i}{\hat{\sigma}}\right| \le k \\ k\ sign\left(\dfrac{e_i}{\hat{\sigma}}\right), & \text{jika } \left|\dfrac{e_i}{\hat{\sigma}}\right| > k \end{cases}$$

Taksiran *robust* yang sering digunakan untuk menaksir σ adalah MADN(e).

Dari penjelasan diatas, maka

$$eff\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) = \frac{\operatorname{var}\left(\hat{\beta}_{jHuber}\right)}{\operatorname{var}\left(\hat{\beta}_{jLS}\right)} = \frac{\left(\frac{MAD(e)}{0.6745}\right)^2 \frac{ave_i\left(\psi\left(\frac{e_i}{\hat{\sigma}}\right)^2\right)}{\left(ave_i\psi'\left(\frac{e_i}{\hat{\sigma}}\right)\right)^2}\frac{n}{n-p}C_{jj}}{MSE\ C_{jj}}$$

Jika $eff\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) < 1$, maka dapat disimpulkan bahwa taksiran parameter pada model regresi robust dengan fungsi Huber lebih effisien daripada taksiran parameter metode OLS .

BAB IV

CONTOH PENERAPAN

Dalam bab ini akan dibahas effisiensi taksiran parameter yang diperoleh dengan metode *Ordinary Least Square* (OLS) dan metode regresi robust dengan fungsi Huber dengan suatu contoh kasus. Ada 2 contoh kasus yang akan diberikan dalam bagian ini, yaitu data yang ada *outlier* dan data yang tanpa *outlier*.

4.1 Kasus Data yang Mengandung *Outlier*

4.1.1 Data

Data yang digunakan mengenai penjualan rumah didaerah Arizona, Amerika Serikat. Data ini terdiri dari 100 observasi dengan variabel yang digunakan harga jual rumah (dalam ribuan dollar), usia rumah (dalam tahun) dan luas tanah (dalam meter persegi). Data penjualan rumah ini memiliki *outlier*, sehingga digunakan sebagai contoh pada tugas akhir ini. Diasumsikan *outlier* yang ada bukan karena suatu kesalahan dalam pengambilan sampel, sehingga *outlier* tidak akan dikeluarkan dari analisis.

4.1.2 Analisis Data

Berdasarkan data yang ada, akan dilihat hubungan antara harga jual rumah dengan usia rumah dan luas tanah. Dalam hal ini, harga jual rumah merupakan variabel dependen, usia rumah dan luas tanah merupakan variabel independen.

Model secara umum dari data yang ada ialah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

dimana y_i = harga jual rumah observasi ke-i (dalam ribuan dollar)

β_j = parameter regresi, j = 0,1,2

x_{i1} = usia rumah observasi ke-i (dalam tahun)

x_{i2} = luas tanah observasi ke-i (dalam meter persegi)

ε_i = *random error* ke-i

Sebelum mencari taksiran parameter, akan diperiksa terlebih dahulu keberadaan *outlier* dalam data. *Outlier* akan dideteksi berdasarkan ukuran *outlier*, seperti *standardized residuals*, *studentized residuals*, *R-student*, *leverage*, DFFITS, DFBETAS, dan *cook's distance* untuk setiap observasi. Pengolahan data dilakukan dengan menggunakan perangkat lunak SPSS 16.0. Pada kasus ini, karena n = 100 dan p = 3, dimana p merupakan banyaknya parameter regresi, maka observasi dikatakan sebagai *outlier* jika nilai

- *standardized residuals* , $|d_i| > 2$

- *studentized residuals*, $|r_i| > 2$
- *R-student*, $|t_i| > 2$
- *Leverage*, $h_{ii} > \frac{2p}{n} = \frac{2.3}{100} = 0.06$
- $|DFFITS_i| > 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{3}{100}} = 0.3464$
- $|DFBETAS_i| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$

Berdasarkan Tabel deteksi *outlier* untuk data penjualan rumah didaerah Arizona, Amerika Serikat saat data ada *outlier* (Tabel 7) terlihat ada beberapa observasi yang memenuhi kriteria sebagai *outlier*. Observasi ke-15, dan 57 diindikasikan kuat merupakan *outlier*.

➢ Pada observasi ke-15,

  - $|d_{15}| = 2.981 > 2$
  - $|r_{15}| = 3.129 > 2$
  - $|t_{15}| = 3.283 > 2$
  - $h_{ii} = 0.082 > 0.06$
  - $|DFFITS_{15}| = 1.047 > 0.3464$
  - $|DFBETAS_{0,15}| = 0.522 > 0.2$
  - $|DFBETAS_{3,15}| = 0.976 > 0.2$

- Pada observasi ke-57,
  - $\left|d_{57}\right| = 2.576 > 2$
  - $\left|r_{57}\right| = 2.662 > 2$
  - $\left|t_{57}\right| = 2.751 > 2$
  - $\left|DFFITS_{57}\right| = 0.718 > 0.3464$
  - $\left|DFBETAS_{1,57}\right| = 0.580 > 0.2$
  - $\left|DFBETAS_{3,57}\right| = 0.659 > 0.2$

Terlihat bahwa observasi ke-15 dan 57 memenuhi banyak kriteria observasi sebagai *outlier*. Oleh karena itu, observasi ke-15 dan 57 merupakan *outlier* dan akan membuat taksiran parameter yang diperoleh dengan metode OLS menjadi tidak bagus. Selain itu, ada observasi-observasi lainnya yang memenuhi kriteria observasi sebagai *outlier* tetapi hanya 1 atau 2 kriteria, seperti observasi ke-6, nilai $\left|DFBETAS_{0,6}\right| = 0.291 > 0.2$ dan $\left|DFBETAS_{1,6}\right| = 0.207 > 0.2$. Hal ini bukan merupakan indikasi yang kuat untuk menyatakan bahwa observasi ke-6 merupakan *outlier*. Demikian juga dengan observasi ke-9, 24, 25 dan beberapa observasi lainnya.

Pengolahan data untuk mencari taksiran parameter baik untuk metode OLS maupun metode regresi robust dengan menggunakan fungsi Huber diperoleh dengan menggunakan perangkat lunak *S-Plus 2000 Professional Release 2*.

Tabel 1

Nilai taksiran parameter dan standar error dari metode OLS saat data ada *outlier*

Variabel	Nilai Taksiran	Std error
Intercept	59666.3935	5216.3246
usia	-2521.9155	260.9813
Luas	2.4044	0.8995

Tabel 2

Nilai taksiran parameter dan standar error metode regresi robust dengan menggunakan fungsi Huber

Variabel	Nilai Taksiran	Std error
Intercept	58991.2068	5149.7002
usia	-2368.1648	257.6480
Luas	1.7946	0.8880

Taksiran parameter yang diperoleh dari metode OLS dan metode regresi robust dengan menggunakan fungsi Huber cukup berbeda, yaitu $\hat{\beta}_{0LS}$ = 59666.3935 sedangkan $\hat{\beta}_{0Huber}$ = 58991.2068, $\hat{\beta}_{1LS}$ = -2521.9155 sedangkan $\hat{\beta}_{1Huber}$ = -2368.1648, dan $\hat{\beta}_{2LS}$ = 2.4044 sedangkan $\hat{\beta}_{2Huber}$ = 1.7946.

Untuk melihat effisiensi relatif taksiran parameter metode regresi robust dengan menggunakan fungsi Huber terhadap taksiran parameter metode OLS, akan dilihat dari variansi masing-masing parameter. Dari tabel 1 dan tabel 2, dapat dilihat nilai standar error dari masing-masing parameter. Karena standar error merupakan akar variansi dari taksiran, maka dapat diperoleh nilai variansi taksiran masing-masing parameter.

Tabel 3

Nilai variansi taksiran parameter dari metode OLS dan metode regresi robust dengan menggunakan fungsi Huber saat data ada *outlier*

Variabel	var OLS	var reg robust
Intercept	27210042.33	26519412.15
usia	68111.23895	66382.49191
Luas	0.80910025	0.788544

Dengan demikian dapat diperoleh nilai effisiensi untuk masing-masing parameter, sebagai berikut

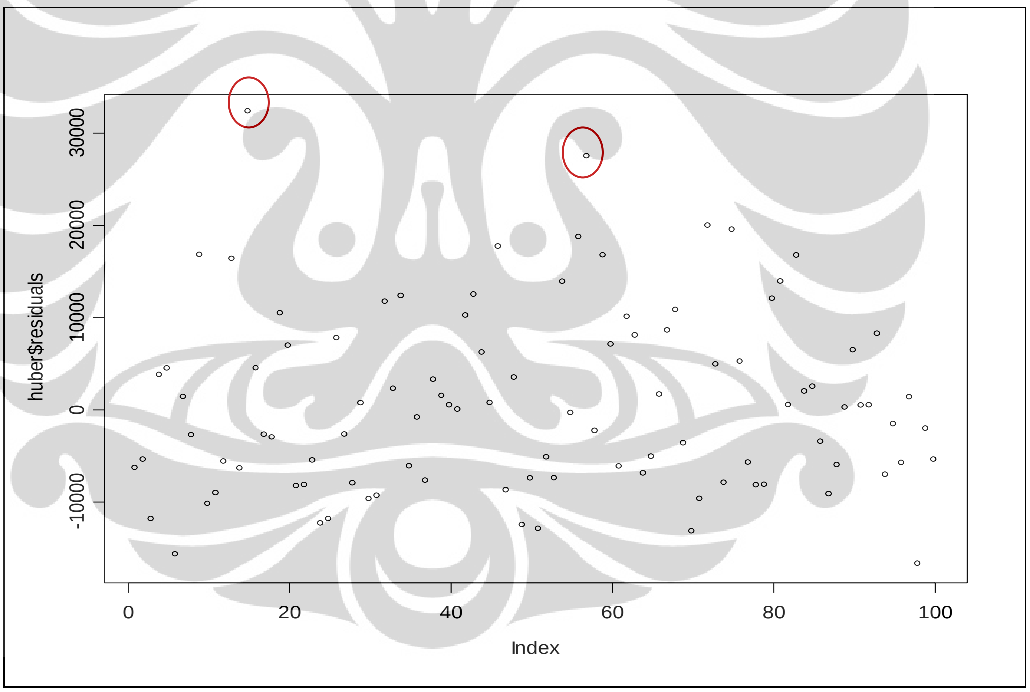
$$eff\left(\frac{\hat{\beta}_{0Huber}}{\hat{\beta}_{0LS}}\right)=\frac{26519412.15}{27210042.33}=0.97461855<1$$

$$eff\left(\frac{\hat{\beta}_{1Huber}}{\hat{\beta}_{1LS}}\right)=\frac{66382.49191}{68111.23895}=0.97461877<1$$

$$eff\left(\frac{\hat{\beta}_{2Huber}}{\hat{\beta}_{2LS}}\right)=\frac{0.788544}{0.80910025}=0.97459369<1$$

Nilai effisiensi untuk masing-masing parameter diperoleh lebih kecil dari 1. Dengan perkataan lain, variansi taksiran parameter metode regresi robust dengan menggunakan fungsi Huber lebih kecil dari taksiran parameter dengan metode OLS. Hal ini menunjukkan bahwa pada contoh diatas, metode regresi robust dengan menggunakan fungsi Huber lebih effisien daripada metode OLS.

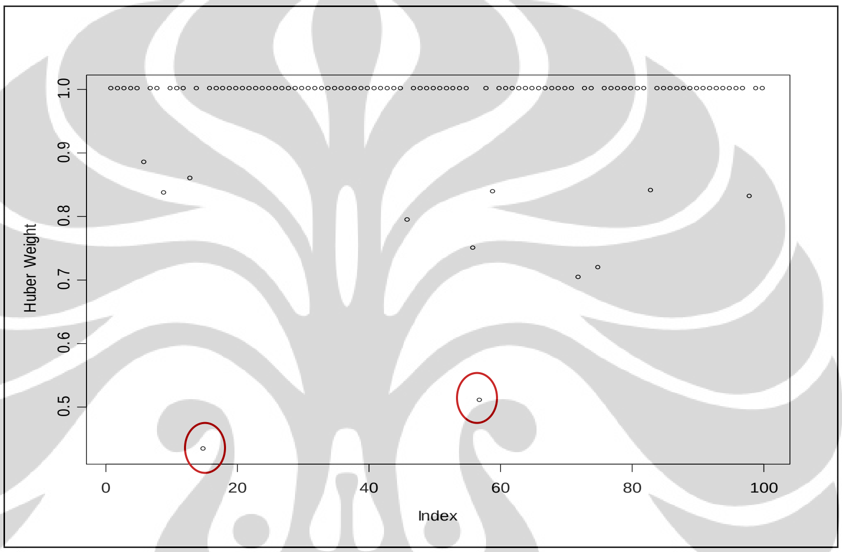
Dari model Huber dapat digambarkan plot *residual* sebagai berikut :



Gambar 2. Plot *residual* metode regresi robust dengan menggunakan fungsi Huber dari setiap observasi

Dari Gambar 2 dapat dilihat bahwa observasi ke-15 memiliki *residual* yang paling besar dibanding observasi lainnya. Selain itu, observasi ke-57 juga memiliki *residual* yang cukup besar.

Bobot yang diberikan untuk setiap observasi dapat digambarkan dalam plot dibawah ini :



Gambar 3. Plot besarnya bobot yang diberikan pada setiap observasi

Dari Gambar 3 terlihat bahwa bobot untuk obervasi ke-15 paling kecil dibandingkan observasi lainnya. Setelah itu, bobot ke-2 terkecil diberikan untuk observasi ke-57. Jadi, dapat disimpulkan bahwa dalam regresi robust dengan menggunakan fungsi Huber, untuk observasi yang mempunyai *residual* besar diberikan bobot kecil.

4.2 Kasus Data Tanpa *Outlier*

4.2.1 Data

Data yang digunakan serupa dengan data pada kasus sebelumnya, yaitu mengenai penjualan rumah didaerah Arizona, Amerika Serikat. Akan tetapi data hanya terdiri dari 60 observasi dan data tidak mempunyai *outlier*. Variabel yang digunakan harga jual rumah (dalam ribuan dollar), usia rumah (dalam tahun) dan luas tanah (dalam meter persegi).

4.2.2 Analisis Data

Akan dilihat hubungan antara harga jual rumah dengan usia rumah dan luas tanah, dimana harga jual rumah merupakan variabel dependen, usia rumah dan luas tanah merupakan variabel independen.

Model secara umum dari data yang ada ialah

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

dimana y_i = harga jual rumah observasi ke-i (dalam ribuan dollar)

β_j = parameter regresi, j = 0,1,2

x_{i1} = usia rumah observasi ke-i (dalam tahun)

x_{i2} = luas tanah observasi ke-i (dalam meter persegi)

ε_i = *random error* ke-i

Sebelum mencari taksiran parameter, akan diperiksa terlebih dahulu keberadaan *outlier* dalam data. Pengolahan data dilakukan dengan menggunakan perangkat lunak SPSS 16.0. Pada kasus ini, karena n = 60 dan p = 3, dimana p merupakan banyaknya parameter regresi, maka observasi dikatakan sebagai *outlier* jika nilai

- *standardized residuals* , $|d_i| > 2$
- *studentized residuals*, $|r_i| > 2$
- *R-student*, $|t_i| > 2$
- *Leverage*, $h_{ii} > \frac{2p}{n} = \frac{2.3}{60} = 0.1$
- $|DFFITS_i| > 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{3}{60}} = 0.4472$
- $|DFBETAS_i| > \frac{2}{\sqrt{n}} = \frac{2}{\sqrt{60}} = 0.2582$

Berdasarkan Tabel deteksi *outlier* untuk data penjualan rumah didaerah Arizona, Amerika Serikat saat data tanpa *outlier* (Tabel 8), terlihat tidak ada data yang mengindikasikan kuat sebagai *outlier*.

Pengolahan data untuk mencari taksiran parameter baik untuk metode OLS maupun metode regresi robust dengan menggunakan fungsi Huber diperoleh dengan menggunakan perangkat lunak *S-Plus 2000 Professional Release 2*.

Tabel 4

Nilai taksiran parameter dan standar error dari
metode OLS saat data tanpa outlier

Variabel	Nilai Taksiran	Std error
Intercept	34048.6779	2390.0428
usia	-548.7666	136.9960
Luas	-0.1937	0.3565

Tabel 5

Nilai taksiran parameter dan standar error
metode regresi robust dengan menggunakan
fungsi Huber saat data tanpa outlier

Variabel	Nilai Taksiran	Std error
Intercept	34202.1236	2655.3015
usia	-558.6715	152.2005
Luas	-0.2016	0.3961

Taksiran parameter yang diperoleh dari metode OLS dan metode regresi robust dengan menggunakan fungsi Huber tidak terlalu berbeda, yaitu $\hat{\beta}_{0LS}$ = 34048.6779 sedangkan $\hat{\beta}_{0Huber}$ = 34202.1236 , $\hat{\beta}_{1LS}$ = -548.7666 sedangkan $\hat{\beta}_{1Huber}$ = -558.6715 , dan $\hat{\beta}_{2LS}$ = -0.1937 sedangkan $\hat{\beta}_{2Huber}$ = -0.2016.

Dari tabel 4 dan tabel 5, dapat diperoleh nilai variansi dari taksiran masing-masing parameter.

Tabel 6

Nilai variansi taksiran parameter dari metode OLS dan metode regresi robust dengan menggunakan fungsi Huber saat data tanpa outlier

Variabel	var OLS	var reg robust
Intercept	5712304.586	7050626.056
usia	18767.90402	23164.9922
Luas	0.12709225	0.15689521

Dengan demikian dapat diperoleh nilai effisiensi untuk masing-masing parameter, sebagai berikut :

$$eff\left(\frac{\hat{\beta}_{0Huber}}{\hat{\beta}_{0LS}}\right)=\frac{7050626.056}{5712304.586}=1.23428>1$$

$$eff\left(\frac{\hat{\beta}_{1Huber}}{\hat{\beta}_{1LS}}\right)=\frac{23164.9922}{18767.90402}=1.23428>1$$

$$eff\left(\frac{\hat{\beta}_{2Huber}}{\hat{\beta}_{2LS}}\right)=\frac{0.15689521}{0.12709225}=1.23449>1$$

Nilai effisiensi untuk masing-masing parameter diperoleh lebih besar dari 1. Dengan perkataan lain, variansi taksiran parameter metode regresi robust dengan menggunakan fungsi Huber lebih besar dari taksiran parameter dengan metode OLS. Hal ini menunjukkan bahwa pada contoh diatas, metode OLS lebih effisien daripada metode regresi robust dengan menggunakan fungsi Huber saat data tanpa *outlier*.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan yang dapat diperoleh dari penulisan skripsi ini adalah

1. Dalam analisis data, saat data yang ada mempunyai *outlier*, taksiran parameter yang *robust* terhadap *outlier* salah satunya dapat diperoleh dengan menggunakan metode regresi robust dengan fungsi Huber.
2. Taksiran parameter untuk $\hat{\beta}_0$ dan $\hat{\beta}_1$ pada model regresi robust sederhana dengan menggunakan fungsi Huber, diperoleh sebagai berikut :

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i w_i y_i - \sum_{i=1}^{n} y_i w_i \sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} x_i^2 w_i - \left(\sum_{i=1}^{n} x_i w_i \right)^2}$$

dengan $w_i = \begin{cases} 1 & \text{jika } |e_i| \le k \\ \dfrac{k}{|e_i|} & \text{jika } |e_i| > k \end{cases}$

Jika diambil nilai terstandardisasi dari e, maka berdasarkan simulasi yang dilakukan oleh Huber dipilih nilai k = 1.345.

3. Taksiran parameter untuk $\hat{\boldsymbol{\beta}}$ pada model regresi robust berganda dengan fungsi Huber, diperoleh sebagai berikut :

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$

4. Efisiensi relatif $\hat{\beta}_{jHuber}$ terhadap $\hat{\beta}_{jLS}$ dapat diukur dengan

$$eff\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) = \frac{\mathrm{var}\left(\hat{\beta}_{jHuber}\right)}{\mathrm{var}\left(\hat{\beta}_{jLS}\right)} = \frac{\left(\frac{MAD(e)}{0.6745}\right)^2 \frac{ave_i\left(\psi\left(\frac{e_i}{\sigma}\right)^2\right)}{\left(ave_i\psi'\left(\frac{e_i}{\sigma}\right)\right)^2}\frac{n}{n-p}C_{jj}}{MSE\ C_{jj}}$$

Jika $eff\left(\frac{\hat{\beta}_{jHuber}}{\hat{\beta}_{jLS}}\right) < 1$, dapat disimpulkan bahwa taksiran parameter pada model regresi robust dengan fungsi Huber lebih effisien daripada taksiran parameter metode OLS .

5.2 Saran

Saran untuk pengembangan skripsi ini adalah

1. Dapat dibahas mengenai pengujian model secara keseluruhan dan pengujian signifikansi dari masing-masing parameter untuk mencari model regresi robust terbaik.
2. Dapat dibahas mengenai metode regresi robust lainnya selain dengan fungsi Huber.

DAFTAR PUSTAKA

Belsley D.A., E. Kuh. & R. Welsch. 1980. *Regression Diagnostics*, hlm. 6-51. John Wiley and Sons, Inc., New York.

Draper, N.H & H. Smith. 1980. *Applied Regression Analysis*, 2nd edition. John Wiley and Sons, Inc., New York.

Huber, P.J. 1981. *Robust Statistics*. New Yoark : John Wiley and Sons.

Maronna, Martin. & Yohai. 2006. *Robust Statistics : Theory and Methods*, xix + 403 hlm. John Wiley and Sons, Ltd., England.

Mendenhall, William. & T. Sincich. 1996*. A Second Course in Statistics Regression Analysis*, 5th edition. xxi + 889 hlm. Prentice Hall, Inc., New Jersey.

Myers, Raymond. 1990*. Classical and Modern Regression with Applications*, 2nd edition. viii + 488 hlm. PWS-KENT Publishing Company, Boston.

Montgomery, Douglas. C, Peck,Elizabeth. A, & Vining, G.G. 2001. *Introduction to Linier Regression Analysis, 3rd edition.* xvi + 641 hlm. John Wiley and Sons, Inc., New York.

Rousseeuw, R.J. & A. M. Leroy. 1987. *Robust Regression and Outlier Detection.* Xiv + 329 hlm. John Wiley and Sons, Inc., New York.

Sugiarti, Harmi. 2008. *Resistensi dan Effisiensi Fungsi Pembobot Huber pada Metode Regresi Robust*.

Wang, G. Y., Xu Lin, Min Zhu, & Z. Bai. 2007. *Robust Estimation Using the Huber Function with a Data-Dependent Tuning Constant*. Journal of Computational and Graphical Statistics, Vol. 16, No. 2, 468-481.

Claerbout, Jon. 1997. *Conjugate-Direction Huber Regression*. http://sepwww.stanford.edu/public/docs/sep92/jon2/paper_htm/index.html , 10 Februari 2009, pk. 20.43.

Jacoby, Bill. 2005. *Regression III: Advanced Methods*. http://polisci.msu.edu/jacoby/icpsr/regress3 , 26 Oktober 2008. pk. 09.06.

Kelly, Gabrielle. 1992. *Robust Regression Estimators-the Choice of Tuning Constants.* http://www.ucd.ie/research/people/mathematicalsciences/drgabrielleekelly/publications , 31 Maret 2009, pk. 18.35.

LAMPIRAN

Lampiran 1

Membuktikan taksiran parameter pada metode regresi linier sederhana dengan metode OLS

Adib : $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ dan $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$.

Bukti :

- Untuk mendapatkan $\hat{\beta}_0$, selesaikan persamaan $n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ (persamaan (2.1.5))

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$n\hat{\beta}_0 = \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^{n} x_i}{n}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Untuk mendapatkan $\hat{\beta}_1$, selesaikan persamaan

  $\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} y_i x_i$ (persamaan (2.1.6)).

  Substitusikan $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ kedalam persamaan 2.1.6, sehingga diperoleh

$$\left(\bar{y}-\hat{\beta}_1\bar{x}\right)\sum_{i=1}^{n}x_i+\hat{\beta}_1\sum_{i=1}^{n}x_i^{\,2}=\sum_{i=1}^{n}y_ix_i$$

$$\bar{y}\sum_{i=1}^{n}x_i-\hat{\beta}_1\bar{x}\sum_{i=1}^{n}x_i+\hat{\beta}_1\sum_{i=1}^{n}x_i^{\,2}=\sum_{i=1}^{n}y_ix_i$$

$$\bar{y}\sum_{i=1}^{n}x_i+\hat{\beta}_1\left(\sum_{i=1}^{n}x_i^{\,2}-\bar{x}\sum_{i=1}^{n}x_i\right)=\sum_{i=1}^{n}y_ix_i$$

$$\hat{\beta}_1\left(\sum_{i=1}^{n}x_i^{\,2}-\bar{x}\sum_{i=1}^{n}x_i\right)=\sum_{i=1}^{n}y_ix_i-\bar{y}\sum_{i=1}^{n}x_i$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}y_ix_i-\bar{y}\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^{\,2}-\bar{x}\sum_{i=1}^{n}x_i}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}y_ix_i-\frac{\sum_{i=1}^{n}y_i}{n}\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^{\,2}-\frac{\sum_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}x_i}$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}y_ix_i-\frac{\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i}{n}}{\sum_{i=1}^{n}x_i^{\,2}-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}$$

$$\hat{\beta}_1=\frac{S_{xy}}{S_{xx}}$$

dimana $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$ dan $\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}$.

$\therefore$ Terbukti bahwa $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$ dan $\hat{\beta}_1=\frac{S_{xy}}{S_{xx}}$

Lampiran 2

Data yang digunakan pada contoh penerapan 4.1, yaitu saat data mengandung outlier

observasi	harga	usia	luas
1	27450	12	1835
2	29850	12	2672
3	25350	11	2430
4	15750	21	1578
5	19650	20	2034
6	23560	10	2246
7	27000	16	2564
8	24000	15	1887
9	65840	7	3754
10	20850	13	1652
11	30000	10	2130
12	28500	12	2031
13	65000	8	4853
14	30150	11	2046
15	74521	12	6524
16	24150	18	1890
17	24450	16	3439
18	21600	17	3329
19	27900	20	3302
20	29100	18	3288
21	22650	15	4209
22	20850	16	4464

23	22950	16	4154
24	30600	11	5621
25	20400	15	4955
26	23850	21	3862
27	22800	18	5140
28	20700	15	2962
29	21300	18	2408
30	24300	13	3282
31	19650	15	3139
32	60000	8	4653
33	30300	15	2600
34	61250	7	3676
35	36000	9	2542
36	25200	15	1498
37	16200	16	1613
38	22800	18	1834
39	43500	9	2463
40	30300	14	2278
41	31950	13	2126
42	35250	16	2254
43	37800	16	2408
44	31200	16	2226
45	29400	16	4274
46	70000	7	5572
47	33900	10	4135
48	27150	18	4129
49	22200	13	3672

50	31350	12	4633
51	25640	12	4502
52	33300	12	4456
53	26250	14	4422
54	31950	19	2330
55	30000	15	3892
56	61290	10	4098
57	80546	5	3351
58	30750	14	4077
59	33540	20	2933
60	34950	15	2511
61	40350	8	3654
62	30270	18	2189
63	26250	19	2395
64	32400	10	2288
65	20400	16	2506
66	24150	17	2164
67	20120	22	2642
68	29700	20	4100
69	21600	17	3672
70	24450	12	3988
71	28050	12	4042
72	70560	7	4624
73	49000	10	4952
74	16350	18	4464
75	68520	8	5042
76	38560	14	4237

77	28500	14	4735
78	28800	13	4951
79	24450	14	3821
80	43650	14	3278
81	61500	7	2945
82	31650	14	3021
83	61875	8	2904
84	21750	18	1950
85	22500	20	4723
86	21600	18	4906
87	34410	10	4654
88	20700	17	4499
89	47550	7	2789
90	33900	15	2279
91	23400	17	2401
92	32850	13	2379
93	55750	7	2895
94	25200	13	2310
95	26250	15	2472
96	26400	13	2265
97	39150	12	4082
98	21450	12	4278
99	29400	13	1867
100	29850	12	2672

Lampiran 3

Output S-PLUS 2000 Professional Release 2 dari contoh penerapan 4.1, yaitu saat data mengandung outlier

```
Working data will be in C:\Documents and
Settings\Vani\SPLUS_Home\_Data
> rumah_read.table("f:\\harga.txt", header=T, row.names=NULL)
> attach(rumah)
> regresi<-lm(harga~usia+luas,rumah)
> summary(regresi)

Call: lm(formula = harga ~ usia + luas, data = rumah)

Residuals:
   Min    1Q Median   3Q   Max
 -18239 -7320 -704.2 6813 29431

Coefficients:
```

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	59666.3935	5216.3246	11.4384	0.0000
usia	-2521.9155	260.9813	-9.6632	0.0000
luas	2.4044	0.8995	2.6730	0.0088

Residual standard error: 9872 on 97 degrees of freedom

Multiple R-Squared: 0.5447

F-statistic: 58.02 on 2 and 97 degrees of freedom, the p-value is 0

```
> library(MASS)

> huber<-rlm(harga~usia+luas,rumah)
> summary(huber)

Call: rlm.formula(formula = harga ~ usia + luas, data = rumah)

Residuals:
   Min     1Q Median   3Q   Max
 -16800 -7049 -261.8 6870 32240
```

Coefficients:

	Value	Std. Error	t value
(Intercept)	58991.2068	5149.7002	11.4553
usia	-2368.1648	257.6480	-9.1915
luas	1.7946	0.8880	2.0208

Residual standard error: 10370 on 97 degrees of freedom

Lampiran 4

Data yang digunakan pada contoh penerapan 4.2, yaitu saat data mengandung outlier

observasi	harga	usia	luas
1	27450	12	1835
2	29850	12	2672
3	25350	11	2430
4	19650	20	2034
5	23560	10	2246
6	27000	16	2564
7	24000	15	1887
8	24550	13	1652
9	30000	10	2130
10	28500	12	2031
11	30150	11	2046
12	24150	18	1890
13	24450	16	3439
14	21600	17	3329
15	29100	18	3288
16	22650	15	4209
17	20850	16	4464
18	22950	16	4154
19	30600	11	5621
20	20400	15	4955
21	23850	21	3862
22	22800	18	5140

23	20700	15	2962
24	21300	18	2408
25	24300	13	3282
26	19650	15	3139
27	30300	15	2600
28	25200	15	1498
29	22800	18	1834
30	30300	14	2278
31	29400	16	4274
32	27150	18	4129
33	22200	13	3672
34	25640	12	4502
35	26250	14	4422
36	30000	15	3892
37	30750	14	4077
38	27270	18	2189
39	26250	19	2395
40	20400	16	2506
41	24150	17	2164
42	20120	22	2642
43	24700	20	2230
44	21600	17	3672
45	24450	12	3988
46	28050	12	4042
47	28500	14	4735
48	28800	13	4951
49	24450	14	3821

50	21750	18	1950
51	22500	20	4723
52	21600	18	4906
53	20700	17	4499
54	23400	17	2401
55	25200	13	2310
56	26250	15	2472
57	26400	13	2265
58	21450	12	4278
59	29400	13	1867
60	29850	12	2672

Lampiran 5

Output S-PLUS 2000 Professional Release 2 dari contoh penerapan 4.2, yaitu saat data tanpa outlier

Working data will be in C:\Documents and Settings\Rianti1\SPLUS_Home_Data

```
> rumah_read.table("f:\\tanpa outlier.txt", header=T, row.names=NULL)
> attach(rumah)
> regresi<-lm(harga~usia+luas,rumah)
> summary(regresi)
```

Call: lm(formula = harga ~ usia + luas, data = rumah)

Residuals:

Min	1Q	Median	3Q	Max
-5559	-2204	-239.7	2612	5566

Coefficients:

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	34048.6779	2390.0428	14.2461	0.0000
usia	-548.7666	136.9960	-4.0057	0.0002
luas	-0.1937	0.3565	-0.5433	0.5890

Residual standard error: 3015 on 57 degrees of freedom

Multiple R-Squared: 0.2232

F-statistic: 8.189 on 2 and 57 degrees of freedom, the p-value is 0.0007481

```
> library(MASS)
> huber<-rlm(harga~usia+luas,rumah)
> summary(huber)
```

Call: rlm.formula(formula = harga ~ usia + luas, data = rumah)

Residuals:

Min	1Q	Median	3Q	Max
-5539	-2224	-214.9	2589	5617

Coefficients:

	Value	Std. Error	t value
(Intercept)	34202.1236	2655.3015	12.8807
usia	-558.6715	152.2005	-3.6706
luas	-0.2016	0.3961	-0.5090

Residual standard error: 3393 on 57 degrees of freedom

Tabel 7

Deteksi *outlier* untuk data penjualan rumah didaerah Arizona, Amerika Serikat saat data ada *outlier*

no	y	x_1	x_2	d_i	r_i	$e_{(i)}$	t_i	h_{ii}	$DFFITS_i$	$DFBETAS_{0,i}$	$DFBETAS_{1,i}$	$DFBETAS_{2,i}$	D_i
1	27450	12	1835	-0.645	-0.656	-6582.213	-0.654	0.023	-0.121	-0.102	0.050	0.096	0.005
2	29850	12	2672	-0.606	-0.611	-6079.935	-0.609	0.007	-0.079	-0.061	0.037	0.042	0.002
3	25350	11	2430	-1.258	-1.273	-12727.664	-1.278	0.014	-0.202	-0.176	0.117	0.123	0.013
4	15750	21	1578	0.532	0.548	5577.398	0.546	0.049	0.136	-0.013	0.089	-0.067	0.006
5	19650	20	2034	0.560	0.573	5777.035	0.571	0.033	0.120	-0.018	0.082	-0.048	0.005
6	23560	10	2246	-1.650	-1.678	-16851.517	-1.694	0.023	-0.315	$\boxed{-0.291}$	$\boxed{0.207}$	0.200	0.033
7	27000	16	2564	0.154	0.155	1544.612	0.154	0.006	0.020	0.003	0.007	-0.009	0.000
8	24000	15	1887	-0.241	-0.244	-2438.433	-0.243	0.016	-0.040	-0.021	-0.001	0.030	0.001
9	65840	7	3754	1.499	1.531	15433.034	1.542	0.031	0.319	$\boxed{0.213}$	$\boxed{-0.270}$	0.006	0.033
10	20850	13	1652	-1.013	-1.031	-10361.554	-1.032	0.025	-0.195	-0.151	0.055	0.163	0.013

Tabel 7 *(lanjutan)*

11	30000	10	2130	-0.969	-0.987	-9926.613	-0.987	0.026	-0.191	-0.178	0.123	0.129	0.012
12	28500	12	2031	-0.586	-0.595	-5954.059	-0.593	0.018	-0.101	-0.084	0.043	0.076	0.003
13	65000	8	4853	1.402	1.434	14483.992	1.442	0.034	0.311	0.063	-0.184	0.161	0.032
14	30150	11	2046	-0.678	-0.689	-6917.627	-0.687	0.022	-0.125	-0.112	0.068	0.091	0.005
15	74521	12	6524	2.981	3.129	32426.231	3.283	0.082	1.047	-0.522	0.044	0.976	0.332
16	24150	18	1890	0.540	0.549	5516.557	0.547	0.023	0.101	0.012	0.047	-0.059	0.003
17	24450	16	3439	-0.318	-0.320	-3178.414	-0.318	0.004	-0.038	0.012	-0.019	-0.008	0.000
18	21600	17	3329	-0.324	-0.327	-3254.178	-0.325	0.007	-0.043	0.017	-0.028	-0.006	0.001
19	27900	20	3302	1.087	1.108	11142.854	1.109	0.027	0.217	-0.128	0.185	0.037	0.016
20	29100	18	3288	0.701	0.709	7079.862	0.707	0.012	0.107	-0.050	0.079	0.015	0.004
21	22650	15	4209	-0.943	-0.952	-9486.233	-0.951	0.009	-0.132	0.063	-0.047	-0.085	0.006
22	20850	16	4464	-0.932	-0.945	-9454.124	-0.944	0.017	-0.157	0.099	-0.076	-0.112	0.008
23	22950	16	4154	-0.644	-0.651	-6491.007	-0.649	0.011	-0.095	0.054	-0.049	-0.058	0.003

Tabel 7 *(lanjutan)*

24	30600	11	5621	-1.503	-1.545	-15678.707	-1.556	0.043	-0.370	0.114	0.049	-0.313	0.045
25	20400	15	4955	-1.352	-1.377	-13842.441	-1.384	0.025	-0.265	0.160	-0.088	-0.221	0.023
26	23850	21	3862	0.796	0.818	8295.260	0.816	0.043	0.193	-0.143	0.168	0.075	0.012
27	22800	18	5140	-0.388	-0.400	-4065.537	-0.398	0.048	-0.099	0.078	-0.060	-0.078	0.003
28	20700	15	2962	-0.837	-0.842	-8356.411	-0.840	0.002	-0.091	-0.013	-0.021	0.021	0.003
29	21300	18	2408	0.125	0.127	1270.419	0.126	0.015	0.020	-0.002	0.012	-0.008	0.000
30	24300	13	3282	-1.061	-1.066	-10583.639	-1.067	0.000	-0.110	-0.040	0.023	0.007	0.004
31	19650	15	3139	-0.986	-0.992	-9843.834	-0.992	0.001	-0.105	-0.004	-0.028	0.009	0.004
32	60000	8	4653	0.944	0.964	9717.601	0.964	0.031	0.199	0.055	-0.126	0.090	0.013
33	30300	15	2600	0.224	0.226	2243.012	0.224	0.004	0.027	0.009	0.004	-0.013	0.000
34	61250	7	3676	1.053	1.076	10842.523	1.076	0.031	0.222	0.154	-0.190	-0.004	0.016
35	36000	9	2542	-0.717	-0.730	-7335.167	-0.728	0.025	-0.138	-0.127	0.105	0.071	0.006
36	25200	15	1498	-0.024	-0.025	-248.438	-0.025	0.026	-0.005	-0.003	0.000	0.004	0.000

Tabel 7 *(lanjutan)*

37	16200	16	1613	-0.708	-0.721	-7236.959	-0.719	0.024	-0.134	-0.060	-0.019	0.104	0.006
38	22800	18	1834	0.417	0.425	4264.454	0.423	0.024	0.080	0.011	0.036	-0.048	0.002
39	43500	9	2463	0.062	0.063	631.545	0.062	0.026	0.012	0.011	-0.009	-0.007	0.000
40	30300	14	2278	0.047	0.047	471.965	0.047	0.009	0.006	0.004	-0.001	-0.004	0.000
41	31950	13	2126	-0.004	-0.004	-44.297	-0.004	0.013	-0.001	-0.001	0.000	0.001	0.000
42	35250	16	2254	1.065	1.076	10732.807	1.077	0.010	0.155	0.044	0.042	-0.092	0.008
43	37800	16	2408	1.286	1.298	12929.171	1.302	0.008	0.177	0.040	0.054	-0.092	0.010
44	31200	16	2226	0.662	0.669	6670.487	0.667	0.011	0.097	0.028	0.026	-0.058	0.003
45	29400	16	4274	-0.019	-0.020	-196.785	-0.020	0.013	-0.003	0.002	-0.002	-0.002	0.000
46	70000	7	5572	1.478	1.532	15681.121	1.543	0.060	$\boxed{0.422}$	0.028	-0.200	$\boxed{0.271}$	0.059
47	33900	10	4135	-1.063	-1.075	-10732.601	-1.076	0.013	-0.164	-0.051	0.092	-0.059	0.009
48	27150	18	4129	0.299	0.304	3045.661	0.302	0.021	0.054	-0.038	0.039	0.030	0.001
49	22200	13	3672	-1.369	-1.376	-13664.298	-1.383	0.001	-0.148	-0.017	0.020	-0.040	0.007

Tabel 7 *(lanjutan)*

50	31350	12	4633	-0.931	-0.943	-9423.604	-0.942	0.014	-0.149	0.027	0.022	-0.106	0.007
51	25640	12	4502	-1.478	-1.494	-14915.625	-1.504	0.012	-0.225	0.031	0.038	-0.150	0.017
52	33300	12	4456	-0.691	-0.698	-6964.770	-0.696	0.011	-0.102	0.012	0.018	-0.067	0.004
53	26250	14	4422	-0.886	-0.895	-8925.208	-0.894	0.011	-0.129	0.052	-0.024	-0.093	0.006
54	31950	19	2330	1.479	1.503	15077.345	1.513	0.022	0.274	-0.044	0.183	-0.094	0.025
55	30000	15	3892	-0.121	-0.122	-1213.266	-0.121	0.005	-0.015	0.005	-0.005	-0.007	0.000
56	61290	10	4098	1.721	1.740	17376.870	1.759	0.012	0.266	0.087	-0.152	0.091	0.023
57	80546	5	3351	2.576	2.662	27164.146	2.751	0.054	0.718	0.580	-0.659	-0.125	0.161
58	30750	14	4077	-0.346	-0.348	-3464.625	-0.347	0.005	-0.043	0.012	-0.007	-0.025	0.001
59	33540	20	2933	1.748	1.780	17899.307	1.801	0.026	0.347	-0.163	0.288	-0.001	0.039
60	34950	15	2511	0.717	0.722	7185.087	0.720	0.005	0.090	0.033	0.012	-0.048	0.003
61	40350	8	3654	-0.803	-0.816	-8193.219	-0.815	0.023	-0.149	-0.100	0.122	0.000	0.007
62	30270	18	2189	1.087	1.103	11044.979	1.104	0.018	0.188	0.002	0.101	-0.089	0.012

Tabel 7 *(lanjutan)*

63	26250	19	2395	0.885	0.900	9022.242	0.899	0.021	0.161	-0.030	0.110	-0.050	0.009
64	32400	10	2288	-0.765	-0.777	-7802.770	-0.776	0.023	-0.142	-0.131	0.094	0.090	0.007
65	20400	16	2506	-0.501	-0.505	-5026.581	-0.503	0.007	-0.066	-0.012	-0.022	0.031	0.001
66	24150	17	2164	0.218	0.221	2206.884	0.220	0.014	0.035	0.006	0.014	-0.019	0.000
67	20120	22	2642	0.971	0.999	10143.929	0.999	0.045	0.242	-0.116	0.210	-0.017	0.019
68	29700	20	4100	1.075	1.101	11136.326	1.102	0.037	0.245	-0.186	0.202	0.119	0.020
69	21600	17	3672	-0.407	-0.412	-4102.441	-0.410	0.009	-0.058	0.030	-0.038	-0.021	0.001
70	24450	12	3988	-1.473	-1.484	-14761.940	-1.494	0.005	-0.184	-0.020	0.052	-0.078	0.011
71	28050	12	4042	-1.122	-1.130	-11245.477	-1.132	0.005	-0.142	-0.011	0.039	-0.065	0.007
72	70560	7	4624	1.765	1.809	18304.370	1.831	0.038	0.410	0.148	-0.291	0.156	0.055
73	49000	10	4952	0.268	0.273	2745.851	0.272	0.026	0.053	-0.002	-0.019	0.036	0.001
74	16350	18	4464	-0.877	-0.894	-8999.103	-0.893	0.028	-0.178	0.133	-0.121	-0.117	0.011
75	68520	8	5042	1.712	1.756	17767.580	1.775	0.039	0.401	0.055	-0.220	0.230	0.052

Tabel 7 *(lanjutan)*

76	38560	14	4237	0.406	0.410	4083.836	0.408	0.007	0.054	-0.019	0.009	0.035	0.001
77	28500	14	4735	-0.734	-0.744	-7447.039	-0.742	0.017	-0.124	0.059	-0.024	-0.099	0.005
78	28800	13	4951	-1.012	-1.028	-10310.945	-1.028	0.022	-0.186	0.074	-0.010	-0.152	0.011
79	24450	14	3821	-0.921	-0.927	-9210.207	-0.927	0.002	-0.103	0.017	-0.014	-0.045	0.004
80	43650	14	3278	1.156	1.161	11524.337	1.164	0.000	0.117	0.019	0.006	-0.002	0.005
81	61500	7	2945	1.257	1.287	13002.332	1.291	0.036	0.283	$\boxed{0.248}$	$\boxed{-0.247}$	-0.093	0.027
82	31650	14	3021	0.003	0.003	26.973	0.003	0.001	0.000	0.000	0.000	0.000	0.000
83	61875	8	2904	1.560	1.590	15996.478	1.603	0.027	0.315	$\boxed{0.276}$	$\boxed{-0.263}$	-0.112	0.033
84	21750	18	1950	0.283	0.287	2881.706	0.286	0.022	0.052	0.005	0.025	-0.029	0.001
85	22500	20	4723	0.194	0.200	2042.829	0.199	0.052	0.051	-0.042	0.039	0.033	0.001
86	21600	18	4906	-0.453	-0.464	-4703.882	-0.462	0.040	-0.106	0.083	-0.067	-0.080	0.004
87	34410	10	4654	-1.137	-1.155	-11575.290	-1.157	0.020	-0.204	-0.015	0.088	-0.120	0.014
88	20700	17	4499	-0.700	-0.712	-7144.857	-0.710	0.023	-0.131	0.091	-0.077	-0.091	0.006

Tabel 7 *(lanjutan)*

89	47550	7	2789	-0.118	-0.121	-1227.908	-0.121	0.038	-0.027	-0.024	0.023	0.010	0.000
90	33900	15	2279	0.667	0.673	6707.342	0.671	0.009	0.092	0.041	0.008	-0.059	0.003
91	23400	17	2401	0.084	0.085	851.147	0.085	0.011	0.012	0.001	0.006	-0.006	0.000
92	32850	13	2379	0.025	0.025	253.051	0.025	0.008	0.003	0.002	-0.001	-0.002	0.000
93	55750	7	2895	0.686	0.703	7106.858	0.701	0.037	0.155	0.137	-0.135	-0.054	0.008
94	25200	13	2310	-0.733	-0.740	-7379.741	-0.738	0.010	-0.104	-0.075	0.030	0.071	0.004
95	26250	15	2472	-0.155	-0.156	-1555.999	-0.156	0.006	-0.020	-0.007	-0.003	0.011	0.000
96	26400	13	2265	-0.600	-0.607	-6050.427	-0.605	0.010	-0.087	-0.063	0.025	0.061	0.003
97	39150	12	4082	-0.007	-0.007	-69.329	-0.007	0.006	-0.001	0.000	0.000	0.000	0.000
98	21450	12	4278	-1.848	-1.865	-18580.200	-1.889	0.008	-0.258	0.011	0.056	-0.150	0.022
99	29400	13	1867	-0.200	-0.203	-2029.067	-0.202	0.019	-0.035	-0.027	0.010	0.028	0.000
100	29850	12	2672	-0.606	-0.611	-6079.935	-0.609	0.007	-0.079	-0.061	0.037	0.042	0.002

Tabel 8

Deteksi *outlier* untuk data penjualan rumah didaerah Arizona, Amerika Serikat saat data tanpa *outlier*

no	y	x_1	x_2	d_i	r_i	$e_{(i)}$	t_i	h_{ii}	$DFFITS_i$	$DFBETAS_{0,i}$	$DFBETAS_{1,i}$	$DFBETAS_{2,i}$	D_i
1	27450	12	1835	0.113	0.117	364.597	0.116	0.045	0.030	0.026	-0.017	-0.019	0.000
2	29850	12	2672	0.963	0.984	3027.521	0.983	0.024	0.203	0.174	-0.144	-0.059	0.014
3	25350	11	2430	-0.727	-0.750	-2331.501	-0.747	0.043	-0.189	-0.174	0.145	0.067	0.012
4	19650	20	2034	-1.005	-1.050	-3305.737	-1.051	0.067	-0.317	0.116	-0.242	0.150	0.034
5	23560	10	2246	-1.514	-1.582	-4981.701	-1.604	0.067	-0.484	-0.461	0.392	0.181	0.076
6	27000	16	2564	0.739	0.748	2281.537	0.745	0.007	0.115	0.017	0.029	-0.055	0.004
7	24000	15	1887	-0.482	-0.491	-1512.042	-0.488	0.023	-0.100	-0.049	0.003	0.076	0.003
8	24550	13	1652	-0.678	-0.699	-2171.846	-0.696	0.042	-0.173	-0.136	0.070	0.129	0.010
9	30000	10	2130	0.614	0.643	2027.099	0.639	0.070	0.197	0.189	-0.156	-0.081	0.013
10	28500	12	2031	0.474	0.488	1513.728	0.485	0.039	0.117	0.104	-0.071	-0.067	0.005
11	30150	11	2046	0.841	0.872	2724.618	0.870	0.053	0.238	0.223	-0.170	-0.119	0.019

Tabel 8 *(lanjutan)*

12	24150	18	1890	0.114	0.118	365.933	0.117	0.040	0.029	-0.002	0.016	-0.018	0.000
13	24450	16	3439	-0.051	-0.051	-155.266	-0.051	0.002	-0.007	0.001	-0.002	-0.002	0.000
14	21600	17	3329	-0.821	-0.831	-2535.476	-0.829	0.007	-0.130	0.050	-0.070	-0.015	0.006
15	29100	18	3288	1.846	1.878	5758.260	1.922	0.017	0.357	-0.189	0.251	0.024	0.041
16	22650	15	4209	-0.780	-0.793	-2428.873	-0.790	0.015	-0.143	0.024	0.007	-0.098	0.007
17	20850	16	4464	-1.179	-1.204	-3706.584	-1.209	0.025	-0.251	0.102	-0.045	-0.188	0.021
18	22950	16	4154	-0.502	-0.510	-1562.927	-0.507	0.015	-0.091	0.034	-0.019	-0.059	0.003
19	30600	11	5621	1.219	1.313	4260.294	1.321	0.120	0.527	0.071	-0.243	0.414	0.091
20	20400	15	4955	-1.478	-1.526	-4747.183	-1.544	0.044	-0.394	0.113	0.015	-0.335	0.050
21	23850	21	3862	0.688	0.722	2286.497	0.719	0.077	0.231	-0.186	0.200	0.060	0.018
22	22800	18	5140	-0.124	-0.130	-410.947	-0.129	0.070	-0.040	0.027	-0.017	-0.031	0.001
23	20700	15	2962	-1.507	-1.520	-4623.723	-1.538	0.001	-0.204	-0.061	0.011	0.039	0.014
24	21300	18	2408	-0.798	-0.815	-2509.012	-0.812	0.025	-0.169	0.040	-0.107	0.076	0.010
25	24300	13	3282	-0.656	-0.665	-2032.969	-0.662	0.010	-0.109	-0.067	0.066	-0.009	0.004

26	19650	15	3139	-1.844	-1.860	-5653.838	-1.902	0.000	-0.248	-0.057	0.014	0.008	0.020
27	30300	15	2600	1.654	1.672	5095.180	1.699	0.005	0.251	0.102	-0.012	-0.117	0.020
28	25200	15	1498	-0.108	-0.112	-346.446	-0.111	0.039	-0.027	-0.014	0.001	0.023	0.000
29	22800	18	1834	-0.337	-0.347	-1079.025	-0.345	0.042	-0.086	0.006	-0.046	0.057	0.003
30	30300	14	2278	1.451	1.474	4513.619	1.490	0.014	0.265	0.176	-0.079	-0.160	0.023
31	29400	16	4274	1.645	1.674	5138.758	1.702	0.018	0.324	-0.125	0.064	0.225	0.034
32	27150	18	4129	1.253	1.283	3959.938	1.291	0.029	0.282	-0.189	0.169	0.148	0.026
33	22200	13	3672	-1.328	-1.348	-4126.877	-1.358	0.013	-0.238	-0.109	0.136	-0.082	0.019
34	25640	12	4502	-0.316	-0.326	-1014.744	-0.323	0.046	-0.083	-0.024	0.048	-0.053	0.002
35	26250	14	4422	0.246	0.251	772.488	0.249	0.025	0.052	0.000	-0.014	0.038	0.001
36	30000	15	3892	1.637	1.657	5057.621	1.684	0.007	0.264	-0.020	-0.014	0.145	0.022
37	30750	14	4077	1.716	1.743	5338.893	1.776	0.014	0.317	0.031	-0.097	0.193	0.032
38	27270	18	2189	1.169	1.197	3697.219	1.202	0.030	0.267	-0.044	0.160	-0.145	0.024
39	26250	19	2395	1.026	1.055	3274.441	1.056	0.039	0.257	-0.094	0.190	-0.102	0.022

40	20400	16	2506	-1.454	-1.472	-4492.712	-1.487	0.008	-0.235	-0.038	-0.058	0.120	0.018
41	24150	17	2164	-0.050	-0.051	-156.454	-0.050	0.021	-0.010	0.000	-0.004	0.006	0.000
42	20120	22	2642	-0.446	-0.475	-1522.825	-0.471	0.101	-0.172	0.110	-0.156	0.033	0.010
43	24700	20	2230	0.683	0.711	2232.252	0.708	0.061	0.206	-0.086	0.163	-0.084	0.014
44	21600	17	3672	-0.799	-0.810	-2475.182	-0.807	0.010	-0.134	0.064	-0.068	-0.048	0.006
45	24450	12	3988	-0.743	-0.761	-2351.138	-0.759	0.030	-0.168	-0.078	0.112	-0.076	0.009
46	28050	12	4042	0.454	0.466	1438.637	0.462	0.031	0.104	0.046	-0.069	0.049	0.004
47	28500	14	4735	1.012	1.040	3224.151	1.041	0.037	0.248	-0.020	-0.058	0.198	0.020
48	28800	13	4951	0.943	0.979	3060.905	0.978	0.054	0.270	0.008	-0.102	0.214	0.024
49	24450	14	3821	-0.390	-0.395	-1206.459	-0.392	0.009	-0.063	-0.012	0.021	-0.030	0.001
50	21750	18	1950	-0.678	-0.697	-2161.049	-0.694	0.038	-0.167	0.016	-0.093	0.104	0.009
51	22500	20	4723	0.113	0.119	378.451	0.118	0.081	0.039	-0.032	0.027	0.023	0.001
52	21600	18	4906	-0.538	-0.559	-1751.459	-0.555	0.058	-0.158	0.107	-0.073	-0.118	0.008
53	20700	17	4499	-1.044	-1.070	-3306.595	-1.072	0.031	-0.240	0.136	-0.090	-0.171	0.019

54	23400	17	2401	-0.283	-0.288	-882.949	-0.286	0.015	-0.052	0.002	-0.024	0.027	0.001
55	25200	13	2310	-0.420	-0.428	-1315.526	-0.425	0.020	-0.083	-0.067	0.042	0.044	0.002
56	26250	15	2472	0.302	0.306	933.692	0.304	0.007	0.047	0.020	-0.002	-0.026	0.001
57	26400	13	2265	-0.025	-0.026	-78.985	-0.025	0.021	-0.005	-0.004	0.003	0.003	0.000
58	21450	12	4278	-1.720	-1.769	-5484.744	-1.803	0.038	-0.434	-0.157	0.269	-0.244	0.060
59	29400	13	1867	0.944	0.969	2996.917	0.968	0.033	0.222	0.177	-0.097	-0.153	0.016
60	29850	12	2672	0.963	0.984	3027.521	0.983	0.024	0.203	0.174	-0.144	-0.059	0.014